Prediction Models: Novel Ways for External Validation and Comparison of Their Performance

An application to patients with Chronic Obstructive Pulmonary Disease (COPD)

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Beniamino Guerra

aus

Italien

Promotionskommission

Prof. Dr. Milo Alan Puhan (Vorsitz)

Prof. Dr. Leonhard Held

Dr. Gerben ter Riet

Zürich, 2017

ABSTRACT

Tailoring medical decisions to the individual patient while considering the scientific evidence has always been the goal of evidence-based medicine. Similarly, personalized medicine or similar terms refer to the pursuit of health care that considers an individual's predisposition, risk factors or manifest illness but also values and preferences to inform medical decisions. A possible approach for personalized medicine is to stratify patients according to the risk of disease or adverse outcomes as estimated by prediction models that may include any indicators of the factors listed before. Estimated risks help balancing expected benefits and harms of preventive or therapeutic interventions for an individual. Hundreds of prediction models have been developed across the medical field. Beside their clinical application, prediction models are also used in research to control for confounding, to efficiently explore subgroup effects or to identify patients that need to be in a predefined risk group in order to be eligible for randomized controlled trials.

However, prediction models often lack external validation (to ensure generalizability) and comparisons among each other, which was the motivation for this thesis. This thesis focuses on prediction models for patients with Chronic Obstructive Pulmonary Disease (COPD) to predict exacerbations and mortality but had also the wider scope of methods development for externally validating prediction models and concurrently comparing their performance.

In the first paper of this thesis, we report on a systematic review of prediction models to estimate the risk of exacerbations in patients with COPD. We included 25 studies reporting on 27 models to predict exacerbations. Outcome definitions, the number and type of predictors, time horizon, statistical methods and measures of prediction model performance were so heterogeneous that it was not possible to identify the most accurate prognostic model for exacerbations. We identified a great need for external validation and comparison of available models in order to inform practice on which models to use. In the second paper, we describe the development of multiple score comparison (MSC) meta-analysis that enables external validation and comparison of multiple prediction models or their simplified versions for prognosis, prognostic scores. We provided a two-step approach, which first performs meta-analyses of the performance of prognostic scores among cohort studies that have data for the same prognostic scores and then aggregates the results across the groups of cohorts. The method builds upon network meta-analytic techniques used for multiple treatment effect comparisons but deals with a number of challenges, like the correlation of data within cohorts (that is much more pronounced than in randomized trials), the selection of the predictive performance metric (so far restricted to the area under the curve or discrimination, respectively) as well as the assessment of consistency, transitivity and heterogeneity. In the third paper, we applied our methodology to patients with COPD. We had at our disposal a large-scale database with 24 international cohort studies and 15762 patients (1871 deaths and 42203 person-years of follow-up) from primary, secondary and tertiary care settings that provided data to calculate between two and nine prognostic scores depending on the cohort study. The ADO score (including age, dyspnea and forced expiratory volume on lung function measurement) outperformed all other scores to predict 3-years mortality in COPD. The assumption of transitivity was not violated. Heterogeneity across direct comparisons was small and we did not identify any substantial local or global inconsistency. Beside the predictive performance, the clinical applicability, namely the availability of predictors and associated cost, and potential harm for patients, should be considered to recommend any

prognostic score for practice. In this respect, the ADO score also appeared most attractive because of its simplicity and easy availability of predictors.

This thesis showed that prognostic scores to estimate the risk for COPD exacerbations are not ready for practice because they lack external validation and comparisons among each other. An accurate prediction model for exacerbations is, however, of great important for personalized care of COPD patients since prognostic knowledge provides great value for the benefit-harm assessment of drug and non-drug therapies and supports the prevention of exacerbation as a key goal of COPD management. The methodology for MSC meta-analysis developed and applied here shows a novel and much needed way to externally validate and compare prognostic scores. While we showed an example of its application for prediction models to estimate the risk of mortality in COPD patients, MSC meta-analysis can be applied to any field of medicine and addresses the great need for external validation and comparison of prediction models. Thereby, the best prediction models can be identified paving the way for risk-stratified, personalized medicine.

ZUSAMMENFASSUNG

Das Ziel evidenzbasierter Medizin ist es seit jeher, medizinische Entscheidungen unter Einbeziehung der wissenschaftlichen Evidenz für den einzelnen Patienten zu finden. Gleichermassen verweisen personalisierte Medizin und ähnliche Begriffe auf eine medizinische Versorgung, welche die individuelle Prädisposition, Risikofaktoren und manifeste Krankheiten, aber auch Wertvorstellungen und Präferenzen bei medizinischen Entscheidungen berücksichtigt. Ein möglicher Ansatz zur personalisierten Medizin ist die Stratifizierung von Patienten nach Krankheitsrisiko oder dem Risiko von Ereignissen. Diese können mittels Prognosemodellen geschätzt werden, die verschiedene der zuvor genannten Faktoren mit einbeziehen. Risikoschätzungen helfen dabei, den erwarteten Nutzen und die Risiken präventiver oder therapeutischer Interventionen gegeneinander abzuwägen. Eine Vielzahl von Prognosemodellen ist im medizinischen Bereich entwickelt worden. Neben ihrer klinischen Anwendung finden sie auch in der Forschung Verwendung, etwa zur Kontrolle von Confounding (Störgrössen), zur effizienten Untersuchung subgruppenspezifischer Effekte oder zur Identifizierung von Patienten, welche, um in eine klinische Studie aufgenommen werden zu können, einer vorspezifizierten Risikogruppe angehören müssen. Jedoch fehlt es oft an einer externen Validierung der Prognosemodelle (um die Generalisierbarkeit sicherzustellen) sowie am Vergleich zu anderen Modellen. Dies stellt die Hauptmotivation für die vorliegende Dissertation dar. Ihr Fokus liegt auf Prognosemodellen für Exazerbation und Mortalität bei Patienten mit chronisch obstruktiver Lungenerkrankung (chronic obstructive pulmonary disease, COPD), sie behandelt aber auch allgemeiner Methoden zur externen Validierung und zum gleichzeitigen Vergleich von Prognosemodellen.

Der erste Artikel dieser Dissertation stellt einen systematischen Review von Prognosemodellen zur Schätzung des Risikos einer Exazerbation bei COPD-Patienten dar. Insgesamt 25 Studien mit 27 Modellen zur Vorhersage einer Exazerbation wurden einbezogen. Die Definition des Outcomes, Zahl und Art der Prädiktoren, Zeithorizonte sowie statistische Methoden und Masse für die Prognosegüte waren so heterogen, dass es nicht möglich war, das genaueste Vorhersagemodell auszumachen. Es wurde deutlich, dass ein grosser Bedarf nach externer Validierung und dem Vergleich verfügbarer Modelle besteht, um zu bestimmen, welche Modelle in der Praxis verwendet werden sollten. Im zweiten Artikel beschreiben wir die Entwicklung der Multiple Score Comparison (MSC) Meta-Analyse, die es ermöglicht, verschiedene Prognosemodelle oder -scores extern zu validieren und zu vergleichen. Wir schlagen ein Zweischrittverfahren vor: Zunächst wird eine Meta-Analyse für die Güte von prognostischen Scores über verschiedene Kohortenstudien, deren Daten die Berechnung der selben Scores erlauben, durchgeführt. Die Ergebnisse werden über die verschiedenen Kohortenstudien aggregiert. Die Methode baut auf Techniken der Netzwerkmetaanalyse zur Kombination mehrerer Vergleiche von Behandlungseffekten auf. Sie muss jedoch mit verschiedenen Schwierigkeiten umgehen können, etwa der Korrelation von Daten aus der selben Kohorte (welche weitaus stärker ausgeprägt ist als bei randomisierten Studien), der Auswahl von Massen für die Prognosegüte (bisher auf die Area under the Curve bzw. Diskriminierung beschränkt) sowie der Beurteilung von Konsistenz und Transitivität. Im dritten Artikel wenden wir unsere Methode auf COPD-Patienten an. Hierfür stand uns eine grosse Datenbank mit 24 internationalen Kohortenstudien und 15762 Patienten (1871 Todesfälle und 42203 Personenjahre Follow-up) aus Primär-, Sekundär- und Tertiärversorgung zur Verfügung. Je nach Kohortenstudie erlaubten diese Daten die Berechnung von zwischen zwei und neun prognostischen Scores. Der ADO Score (welcher Alter, Atemnot

und das forcierte exspiratorische Volumen aus einem Lungenfunktionstest beinhaltet) schnitt bei der Vorhersage von Dreijahresmortalität bei COPD-Patienten am besten ab. Die Transitivitätsannahme war nicht verletzt, die Heterogenität zwischen direkten Vergleichen war gering und wir konnten keine wesentliche lokale oder globale Inkonsistenz feststellen. Neben der Vorhersagegüte sollte die klinische Anwendbarkeit, insbesondere die Verfügbarkeit von Prädiktoren und die damit verbundenen Kosten, sowie potenzielle Risiken für den Patienten bei der Empfehlung eines Scores mit einbezogen werden. Auch in dieser Hinsicht erschien der ADO Score aufgrund seiner Einfachheit und der leichten Verfügbarkeit der Prädiktoren am attraktivsten.

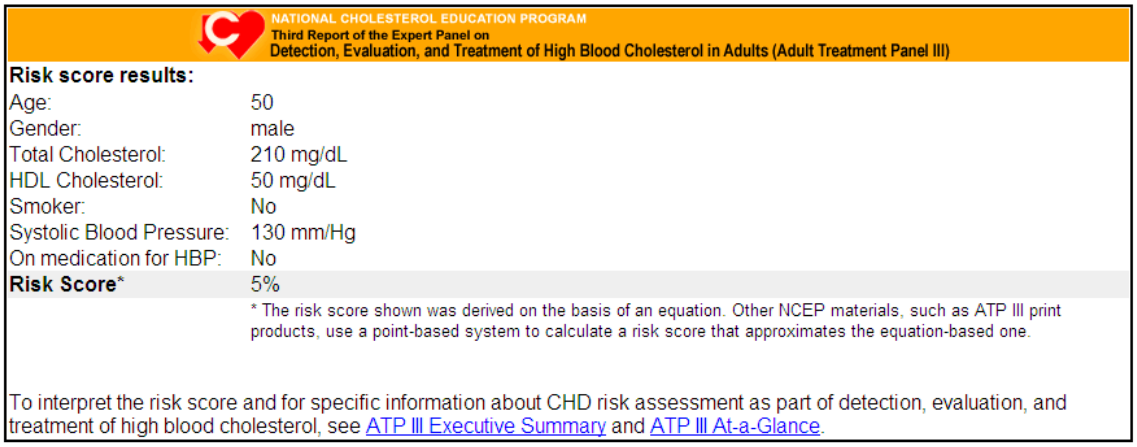
Die vorliegende Dissertation zeigt, dass prognostische Scores zur Schätzung des Risikos einer Exazerbation bei COPD-Patienten für die Praxis noch nicht bereit sind, da externe Validierung und Vergleiche zwischen Scores fehlen. Ein präzises Prognosemodell für Exazerbationen ist jedoch von grosser Wichtigkeit für die individuelle Behandlung von COPD-Patienten. Prognostisches Wissen ist wertvoll für Nutzen-Risiko-Analysen von medikamentösen und nicht-medikamentösen Behandlungen und unterstützt mit der Vermeidung von Exazerbationen eines der Kernziele des COPD-Managements. Die hier entwickelte und angewandte MSC-Metaanalyse-Methode stellt ein neues und dringend benötigtes Verfahren zur externen Validierung und zum gleichzeitigen Vergleich prognostischer Scores dar. Während unser Anwendungsbeispiel Prognosemodelle zur Schätzung des Mortalitätsrisikos von COPD-Patienten behandelt kann die Methode in jedem Teilgebiet der Medizin Anwendung finden und begegnet dem grossen Bedarf zur externen Validierung und zum Vergleich von Prognosemodellen. Auf diese Weise können die besten Prognosemodelle ausgemacht werden und den Weg für risikostratifizierte, personalisierte Medizin bereiten.

CONTENTS

Prediction models in medical practice

Prediction models assess a person's risk of developing a disease or an adverse outcome in the future. They are supposed to be helpful in clinical practice, public health management and medical research[1] and it has been shown that prediction models can be as accurate as physicians' assessments[2,3] and realistically complement them in a fruitful way. In some medical fields, leading guidelines strongly recommend the use of prediction models or the underlying risk prediction scores for clinical practice.[4] For instance, in cardiovascular medicine, risk scores are used in clinical practice to guide primary and secondary prevention of cardiovascular disease.[5–8] Many other clinical fields also know prediction models.[9–15]

The most renowned example of a prognostic model is probably the Framingham risk score to estimate the risk of myocardial infarction or stroke over 10 years.[16] Online, tablet or smartphone applications have been made available (Figure 1) that make it easy for clinicians to use such risk scores. For example, it predicts a 5% risk of a myocardial infarction or stroke over 10 years for a 50-year-old non-smoking male with 210 mg/dL total cholesterol, 50mg/dL High-Density Lipoprotein (HDL) cholesterol, 130 mm/Hg systolic blood pressure and not on medication for high blood pressure Figure 1.



NATIONAL CHOLESTEROL EDUCATION PROGRAM
Third Report of the Expert Panel on
Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III)

Risk score results:

Age:	50
Gender:	male
Total Cholesterol:	210 mg/dL
HDL Cholesterol:	50 mg/dL
Smoker:	No
Systolic Blood Pressure:	130 mm/Hg
On medication for HBP:	No
Risk Score*	5%

* The risk score shown was derived on the basis of an equation. Other NCEP materials, such as ATP III print products, use a point-based system to calculate a risk score that approximates the equation-based one.

To interpret the risk score and for specific information about CHD risk assessment as part of detection, evaluation, and treatment of high blood cholesterol, see ATP III Executive Summary and ATP III At-a-Glance.

Figure 1 Framingham score assessment (https://www.nhlbi.nih.gov/files/docs/guidelines/atglance.pdf)

The estimated risk supports the treating physician in the preventive or therapeutic choice according to the expected benefit-harm balance, that is, to prescribe treatments only if the benefits (in terms of reduced risk of adverse outcome) are expected to outweigh the harms. For example, in cardiovascular primary prevention, the use of lipid lowering drugs in addition to a healthy lifestyle is only recommended if the 10-year risk of a myocardial infarction or stroke exceeds some risk threshold (e.g. 10%) since in lower risk people the harms from these drugs outweigh the expected benefit.

Risk thresholds for recommending for or against treatment can be determined using quantitative benefit-harm assessment that takes into account outcome risks of a population or its subgroups, treatment effects as estimated in randomized trials as well as the importance of benefit and harm outcomes. For example, a quantitative benefit-harm assessment suggested that a novel drug for patients with chronic obstructive pulmonary disease (COPD) is only indicated if the one year risk for very severe exacerbations requiring a hospital admission is at least 20%.[17] Below this risk, the harm from gastrointestinal, psychiatric and neurological side effects is very likely to outweigh the benefits of the novel drug. For clinicians to act on such evidence it is a prerequisite to estimate risks for such events accurately based on prediction models or their simplified version for prognosis, namely, prognostic scores.

Prediction Models in Medical Research

Prediction models can also serve medical research, for instance for the design of randomized controlled trials (RCT), for efficiently identifying subgroup effects, for statistical analysis (adjustment for confounders) in RCTs and observational studies.

RCT should define eligibility in a way that ensures the safety of patients both with respect to benefit and harm outcomes. In some situations, for example as described before for the novel drugs to prevent exacerbations in COPD patients, harms can be expected so that only patients at higher risk for the outcome to be prevented should be included in whom the benefit harm balance may be favorable. In other instances, there may be a specific side effect certain participants are susceptible to and where it is not safe to enroll them into a RCT. For example, people at high risk for gastrointestinal bleeds have been excluded from RCTs on low dose aspirin for primary cardiovascular prevention to minimize risks. In these situations, prediction models can help to explicitly estimate the risk of the outcome of interest and include or exclude patients accordingly.

Another application is pre-stratification in an RCT, in order to control for confounding. Confounders are defined as variables that are associated with both the exposure and the outcome and do not act as mediators, possibly biasing the relationship of exposure and outcome. In small RCTs or even in moderately sized RCTs with populations that have diverse outcome expectations, randomization may not yield balanced groups. Pre-stratification offers additional control for confounding. Prediction models carry more information than single characteristic so that the predicted risk can be used for more efficient pre-stratification. Similar efficiency considerations apply to the statistical control for confounding through adjustment or for the identification of effect modification in RCTs or observational studies. If the number of events is limited statistical models may not afford many covariates to adjust for confounding or to identify differences in effects across subgroups. Prediction models, by combining pertinent outcome predictors, carries more information than single characteristics and offer a way for more parsimonious models without losing (much) information.

The problem of lacking validations and comparisons of prediction models

Most clinical fields have still a long way to go on the path between model development and clinical application since the generalisability of the models and their clinical effectiveness must be ensured. In

particular, the generalisability of a prediction model is ensured by demonstrating good accuracy (i.e. prognostic properties in its derivation cohort, namely the cohort in which it was developed), good reproducibility (i.e. a good performance of the model in the same source population of the derivation cohort) and good transportability (i.e. a good performance of the model in a different population).[18] Researchers often develop new models instead of checking for the generalisability of the existing model and the need for a new model at all first. Steyerberg provided an excellent framework for how to build upon existing models before embarking on model extensions or even the development of new prediction models.[1]

Prediction models have been developed to assess the individual risk of adverse outcomes in various medical fields. For instance, over the past two decades, numerous prediction models have been developed to estimate the risk of developing cardiovascular disease. A recent systematic review concerning prediction models for cardiovascular disease risk in the general population[19] included 212 studies, referring to the development of 363 prediction models and 473 external validations. Only 132 of the developed models (36%) were externally validated, and only 70 (19%) of these were validated by independent investigators.

Even if many scores have been developed and published, they are sometimes not used in clinical practice. One reason for general practitioners not using them is the lack of external validation so that they do not know if a prognostic model predicts a risk accurately for "their" patients[20,21], a prerequisite for generalizability[18,22]. Indeed, a prerequisite for the models' reliability is that the risk of adverse outcome(s) can be accurately predicted by a prediction model[1,23,24] that has been thoroughly developed and validated.[25–28] In addition, Steyerberg's framework for updating prediction models starting with a scale that goes from simply adjusting the intercept to developing (and validating) an entirely new model is rarely followed.[1,29,30] Furthermore, many validation studies still do not validate (or update) the original prognostic score, but rather develop a second model. This practice has led to numerous prognostic scores for the same conditions that have not been externally validated. If prognostic scores are validated at all, there is usually a focus on a single score, while comparisons with other prognostic scores are lacking. The applicability of prediction models is hindered by the lack of comparisons among available prognostic scores.[31] It is not yet clear which prediction model predicts mortality most accurately and is applicable in daily practice.

Need of a New Methodology

Ideally, researchers should perform an external validation and comparison of all the developed scores.[31] Indeed, the collection of "big data" is becoming more common[32] as is the growing availability of individual patient data (IPD).[33–38] International collaborative efforts[39–41] start to give life to the call of the medical community for data sharing.[42] Thus, researchers are provided with new opportunities (and challenges)[43,44]. For example, there may be the possibility of checking a model's predictive performance across clinical settings, populations, and subgroups[43] or of updating (or recalibrating) models,[1] or allowing for head-to-head comparisons;[43,45] a challenge is checking the transportability associated with external validation studies.[46] A possible approach to compare several scores together across different cohorts is to pool available databases and then perform the comparative analysis on the single pooled cohort. A recent example where available databases are pooled together in order to compare all the

available prediction scores together is the COCOMICS study that compared several scores to estimate the risk of mortality in patients with COPD.[47]

Extending the simple-pooling approach to pool direct comparisons taken from different studies is a logical next step, to explore heterogeneity and its effects on the usefulness of prognostic scores. Indeed, in prognostic performance comparison, simply reporting a model's overall performance (averaged across all clusters and individuals) is not sufficient because it can mask differences across these clusters and subgroups (in the same way, as analyzing pooled data not weighting them, would ignore characteristics of the subgroups and individual studies being pooled in comparison to a meta-analysis)[48,49]. Potential users need to know whether or not a model is reliable or transportable to all the settings, populations, and groups represented in the data,[43] relying as well on a careful interpretation of the external validation studies.[46]

Prediction models in COPD

COPD is a complex disease and medical literature indicates management guidelines that categorize patients according to the severity of airflow limitation (GOLD 1, 2, 3, 4) according to the FEV1 % of the predicted value for COPD patients with of a specific sex, age height and ethnicity).[50–53] However, there is weak correlation between this categorization and symptoms of a COPD patient. Thus, a more comprehensive categorization (GOLD A, B, C, D) was created, to have a better understanding of the impact of the disease on COPD patients. This new tool includes, not only the spirometric grade (i.e. airflow limitation), but also the exacerbation risk of the patient, in order to highlight the importance of prevention and quality of life.[50] Both of these assessments were proven to poorly perform for prediction of mortality and other health-related outcomes.[40] Thus, for the purpose of prediction, without considering the symptom assessment, specific prediction models have to be developed.

Indeed, starting in 2004 with the BODE index,[54] several prognostic scores incorporating clinical and non-clinical factors (including symptoms, patient history, functional capacity tests and biomarkers) have been developed to estimate the risk of mortality,[54–67] or exacerbations[60,61,68–95], the course of health-related quality of life[70] or resource utilisation[91] in patients with COPD. For COPD, surveys show that the majority of physicians do not consider the patient's prognosis (i.e. in terms of exacerbations or mortality) when prescribing COPD treatments unless the patients have very severe disease (e.g. requiring oxygen or surgery) and that current medical practice deviates substantially from the GOLD and other guidelines.[39,96–99]

Exacerbations are an ideal target for risk-stratified treatment since they are one of the most important outcomes for COPD patients and avoiding them leads to a higher health-related quality of life, longer life and lower health-care costs. "All-cause mortality" is a more consolidated outcome because of the hard endpoint definition and has a longer history in the medical literature; the first developed score dates back to 2004[54]. Today, no treatments to lower the risk of mortality are yet available for patients with COPD; thus, for this outcome, prediction scores cannot provide risk-stratified treatment guidance. However, prognostic scores may help to make randomized trials with all-cause mortality as a primary outcome more efficient by only including patients at higher risk.[100,101] Furthermore, for such multi-morbid patients, as COPD patients (where cardiovascular disease, diabetes, renal disease and lung cancer can contribute to death)[102,103] high risk of death for COPD has implications on which could be the

optimal prevention and treatment of cardiovascular prevention, lung cancer screening or other treatments. Thus, over- and under-treatment in COPD could be avoided.

Thesis Outline

This thesis focuses on prediction models for patients with Chronic Obstructive Pulmonary Disease (COPD) to predict exacerbations and mortality but had also the wider scope of methods development for externally validating prediction models and concurrently comparing their performance.

In the second chapter, we report on a systematic review of prediction models to estimate the risk of exacerbations in patients with COPD. Chapter 3 describes the development of multiple score comparison (MSC) meta-analysis that enables external validation and comparison of multiple prediction models. In chapter 4, the first application of MSC meta-analysis methodology is described. For that purpose, we used a large-scale database with 24 international cohort studies and 15762 patients that provided data to calculate and compare between two and nine prognostic scores depending on the cohort study. Finally, chapter 5 includes a discussion of findings, their implication for the practice as well as for further research in COPD and beyond.

A brief introduction of the chapters follows:

CHAPTER 1: General Introduction

CHAPTER 2: Prediction models for exacerbations in patients with COPD: A systematic review

This paper illustrates a systematic review of models predicting exacerbations in patients with COPD. We identified 25 studies reporting on 27 statistical prediction models for exacerbation in patients with COPD. The prediction models differ greatly in terms of how they were developed and which predictors and measures for their predictive performance were used. Most studies were of good quality concerning the clinical settings and tests (i.e. selection, definition and measurement of predictors and outcomes and in terms of how patients were selected). However, most of the prediction models were at high risk of bias because of unsound statistical methods used to develop prediction models and a lack of validation. The overall assessment of readiness of the 27 prediction models for use in practice showed that none were ready for clinical application.

CHAPTER 3: Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores

This paper illustrates a novel methodology that we named “Multiple Score Comparison meta-analysis”, that allows one to externally validate different prognostic scores and concurrently compare their

predictive performance. It is a frequentist two-stage network meta-analytic approach to compare all scores within a single analytical framework accounting for correlations between scores within cohorts. We assessed heterogeneity, inconsistency and transitivity and provided a performance ranking of the prognostic scores.

CHAPTER 4: A novel comprehensive approach for large-scale external validation and comparison of prognostic models: An application to chronic obstructive pulmonary disease

This paper refers to the clinical application of the MSC meta-analysis to a large-scale database of patients with COPD. This study was based on 26 cohort studies of the COPD Cohorts Collaborative International Assessment (3CIA) consortium. Our study had 2 main findings. Firstly, our results indicated that the ADO index has the best ability to predict 3-year mortality in patients with COPD, followed by the updated BODE and eBODE indices. Given its simplicity, the ADO index may be the most attractive option across care settings to inform patients and health care professionals about prognosis and to inform treatment decisions whose effectiveness may depend on life expectancy. Secondly, we showed how our novel methodology (MSC meta-analysis) meets the call of top medical journals for new approaches for external validation and concurrent comparison of risk prediction models and scores that should take advantage of data sharing, individual patient data (IPD) and advanced analytical techniques.

CHAPTER 5: General Discussion

References

1. Steyerberg EW. *Clinical Prediction Models*. (Gail M, Krickeberg K, Sarnet J, Tsiatis A, Wong W, eds.). Springer - Statistics for Biology and Health - ISBN: 978-1-4419-2648-7; 2010.

2. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: A meta-analysis. *Psychol Assess*. 2000;12(1):19-30. doi:10.1037//1040-3590.12.1.19.

3. Collins GS. The BMJ Gary Collins : Opening up multivariable prediction models. Bmj.

4. Van Gorp MJ, Steyerberg EW, Van Der Graaf Y. Decision Guidelines for Prophylactic Replacement of Bj??rk-Shiley Convexo-Concave Heart Valves: Impact on Clinical Practice. *Circulation*. 2004;109(17):2092-2096. doi:10.1161/01.CIR.0000125853.51637.C8.

5. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation*. 2014;129(25 SUPPL. 1). doi:10.1161/01.cir.0000437738.63853.7a.

6. World Health Organization. *Prevention of Cardiovascular Disease.*; 2007. doi:10.1001/jama.2011.1668.

7. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke

Association. *Stroke*. 2011;42(2):517-584. doi:10.1161/STR.0b013e3181fcb238.

8. Expert Panel on Detection, Evaluation and T of HBC in A. Executive summary of the third report of the national cholesterol education program Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Jama*. 2001;285(19):2486-2497. doi:10.1001/jama.285.19.2486.

9. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38. doi:10.1186/1472-6947-6-38.

10. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J Cancer*. 1982;45(3):361-366. doi:10.1038/bjc.1982.62.

11. Di M-Y, Liu H, Yang Z-Y, Bonis PAL, Tang J-L, Lau J. Prediction Models of Mortality in Acute Pancreatitis in Adults. *Ann Intern Med*. 2016;(38). doi:10.7326/M16-0650.

12. Herman WH. Predicting risk for diabetes: Choosing (or building) the right model. *Ann Intern Med*. 2009;150(11):812-814. doi:10.7326/0003-4819-150-11-200906020-00010.

13. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission A Systematic Review. *JAMA*. 2015;306(15).

14. Ma D, Watson L, Ja R, et al. Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes (Review) Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes. 2011;(11). doi:10.1002/14651858.CD004902.pub4.Copyright.

15. Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. *BMJ*. 2016;354(19):i4338. doi:10.1136/bmj.i4338.

16. Wilson PWF, D'Agostino RB, Levy D, Belanger a. M, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation*. 1998;97(18):1837-1847. doi:10.1161/01.CIR.97.18.1837.

17. Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. *Thorax*. 2014;69(7):616-622. doi:10.1136/thoraxjnl-2013-204155.

18. Justice AC, Covinsky KE, Berlin J a. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-524. doi:199903160-00009 [pii].

19. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353(i2416):11. doi:10.1136/bmj.i2416.

20. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? *Bmj*. 1995;311(7019):1539-1541. doi:10.1136/bmj.311.7019.1539.

21. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007.

22. Collins GS, de Groot J a, Dutton S, et al. External validation of multivariable prediction models: a

systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40. doi:10.1186/1471-2288-14-40.

23. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research 1: what , why , and how ? *BMJ*. 2009;338(b375):1317-1320. doi:10.1136/bmj.b375.

24. Steyerberg EW, Moons KGM, Windt DA Van Der, et al. Prognosis Research Strategy (PROGRESS) 3 : Prognostic Model Research. *PLoS Med*. 2013;10(2 (e1001381)):9. doi:10.1371/jour-.

25. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247.

26. Bleeker S., Moll H., Steyerberg EW, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol*. 2003;56(9):826-832. doi:10.1016/S0895-4356(03)00207-5.

27. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research 3: validating a prognostic model. *BMJ*. 2009;338(b605):1432-1435. doi:10.1136/bmj.b605.

28. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.

29. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586. doi:10.1002/sim.1844.

30. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method to update prediction models. *Stat Med*. 2014;Under revi(October). doi:10.1002/sim.7179.

31. Collins GS, Moons KGM. Comparing risk prediction models Should be routine when deriving a new model for the same purpose. *BMJ*. 2012;3186(May):1-2. doi:10.1136/bmj.e3186.

32. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):93-101. doi:10.1002/bjs.9723.

33. Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM. Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLoS Med*. 2015;12(10):1-12. doi:10.1371/journal.pmed.1001886.

34. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Assessing risk prediction models using individual participant data from multiple studies. *Am J Epidemiol*. 2014;179(5):621-632. doi:10.1093/aje/kwt298.

35. Debray TP a, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med 2013;32*. 2013;32:3158-3180. doi:10.1002/sim.5732.

36. Broeze KA, Opmeer BC, Bachmann LM, et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol*. 2009;9:22. doi:10.1186/1471-2288-9-22.

37. Ahmed I, Debray TP a, Moons KGM, Riley RD. Developing and validating risk prediction models in

an individual participant data meta-analysis. *BMC Med Res Methodol*. 2014;14(1):3. doi:10.1186/1471-2288-14-3.

38. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. *Br Med J*. 2015;350(1088):1-9. doi:10.1136/bmj.h1088.

39. Siebeling L, ter Riet G, van der Wal WM, et al. ICE COLD ERIC--International collaborative effort on chronic obstructive lung disease: exacerbation risk index cohorts--study protocol for an international COPD cohort study. *BMC Pulm Med*. 2009;9:15. doi:10.1186/1471-2466-9-15.

40. Soriano JB, Lamprecht B, Ramírez AS, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. *Lancet Respir Med*. 2015;3(6):443-450. doi:10.1016/S2213-2600(15)00157-5.

41. Siebeling L, Puhan MA, Muggensturm P, Zoller M, Ter Riet G. Characteristics of Dutch and Swiss primary care COPD patients - baseline data of the ICE COLD ERIC study. *Clin Epidemiol*. 2011;3:273-283. doi:10.2147/CLEP.S24818.

42. Krumholz HM. Why data sharing should be the expected norm. *Bmj*. 2015;350(feb05 6):h599-h599. doi:10.1136/bmj.h599.

43. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *Bmj*. 2016;353(i3140):11. doi:10.1136/bmj.i3140.

44. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *Bmj*. 2012;344(January):1-10. doi:10.1136/bmj.d7762.

45. Moons KGM, Schuit E. Prediction of cardiovascular disease worldwide. *Lancet Diabetes Endocrinol*. 2015;3(5):309-310. doi:10.1016/S2213-8587(15)00002-9.

46. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi:10.1016/j.jclinepi.2014.06.018.

47. Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: The COCOMICS study. *Eur Respir J*. 2013;42(2):323-332. doi:10.1183/09031936.00121012.

48. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol*. 1999;28(1):1-9. doi:10.1093/ije/28.1.1.

49. Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. *Heal (San Fr*. 2001;24(2):218-230.

50. Decramer M, Vogelmeier C, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obrstructive Pulmonary Disease (updated 2017). *Glob Initiat Chronic Obstr Lung Dis*. 2011.

51. Celli BR, MacNee W, Agusti A, et al. Standards for the diagnosis and treatment of patients with

COPD: a summary of the ATS/ERS position paper. *Eur Respir J*. 2004;23(6):932-946. doi:10.1183/09031936.04.00014304.

52. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP.

53. Miravitlles M, Soler-Cataluña JJ, Calle M, et al. Spanish COPD Guidelines (GesEPOC): pharmacological treatment of stable COPD. Spanish Society of Pulmonology and Thoracic Surgery. *Arch Bronconeumol*. 2012;48(7):247-257. doi:10.1016/j.arbres.2012.04.001.

54. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(10):1005-1012. doi:10.1056/NEJMoa021322.

55. Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open*. 2012;2(6):1-10. doi:10.1136/bmjopen-2012-002152.

56. Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet*. 2009;374(9691):704-711. doi:10.1016/S0140-6736(09)61301-5.

57. Soler-Cataluña JJ, Martinez-Garcia MA, Sanchez LS, Tordera MP, Sanchez PR. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. *Respir Med*. 2009;103(5):692-699. doi:http://dx.doi.org/10.1016/j.rmed.2008.12.005.

58. Esteban C, Quintana JM, Aburto M, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. *Respir Med*. 2011;105(11):1662-1670. doi:http://dx.doi.org/10.1016/j.rmed.2011.05.005.

59. Kostianev SS, Hodgev VA, Iluchev DH. Multidimensional system for assessment of COPD patients. Comparison with BODE index. *Folia Medica (Plovdiv)*. 2008;50(4):29-38.

60. Briggs A, Spencer M, Wang H, Mannino D, Sin DD. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. *Arch Intern Med*. 2008;168(1):71-79. doi:http://dx.doi.org/10.1001/archinternmed.2007.37.

61. Schembri S, Anderson W, Morant S, et al. A predictive model of hospitalisation and death from chronic obstructive pulmonary disease. *Respir Med*. 2009;103(10):1461-1467. doi:http://dx.doi.org/10.1016/j.rmed.2009.04.021.

62. Siebeling L, Musoro JZ, Geskus RB, et al. Prediction of COPD-specific health-related quality of life in primary care COPD patients: a prospective cohort study. *npj Prim Care Respir Med*. 2014;24(14060):7. doi:10.1038/npjpcrm.2014.60.

63. Esteban C, Quintana JM, Aburto M, Moraza J, Capelastegui A. A simple score for assessing stable chronic obstructive pulmonary disease. *QJM*. 2006;99(11):751-759. doi:10.1093/qjmed/hcl110.

64. Boeck L, Soriano JB, Brusse-Keizer M, et al. Prognostic assessment in COPD without lung function: The B-AE-D indices. *Eur Respir J*. 2016;47(6):1635-1644. doi:10.1183/13993003.01485-2015.

65. Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. *Chest*. 2005;127(6):1890-1897. doi:10.1378/chest.127.6.1890.

66. Cardoso F, Tufanin AT, Colucci M, Nascimento O, Jardim JR. Replacement of the 6-min walk test with maximal oxygen consumption in the BODE index applied to patients with COPD: An equivalency study. *Chest*. 2007;132(2):477-482. doi:10.1378/chest.07-0435.

67. Williams JEA, Green RH, Warrington V, Steiner MC, Morgan MDL, Singh SJ. Development of the i-BODE: Validation of the incremental shuttle walking test within the BODE index. *Respir Med*. 2012;106(3):390-396. doi:10.1016/j.rmed.2011.09.005.

68. Almagro P, Barreiro B, Ochoa de Echaguen A, et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. *Respiration*. 2006;73(3):311-317. doi:10.1159/000088092.

69. Thomsen M, Ingebrigtsen TS, Marott JL, et al. Inflammatory biomarkers and exacerbations in chronic obstructive pulmonary disease. *JAMA*. 2013;309(22):2353-2361. doi:http://dx.doi.org/10.1001/jama.2013.5732.

70. Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. *Am J Respir Crit Care Med*. 2009;180(12):1189-1195. doi:http://dx.doi.org/10.1164/rccm.200902-0271OC.

71. Müllerova H, Maselli DJ, Locantore N, et al. Hospitalized Exacerbations of COPD. *CHEST J*. 2015;147(4):999-1007. doi:10.1378/chest.14-0655.

72. Almagro P, Soriano JB, Cabrera FJ, et al. Comorbidities and Short-term Prognosis in Patients Hospitalized for Acute Exacerbation of COPD. *Chest*. 2012;142(5):1126-1133.

73. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2010;363(11):1128-1138. doi:10.1056/NEJMoa0909883.

74. Moberg M, Vestbo J, Martinez G, et al. Validation of the i-BODE index as a predictor of hospitalization and mortality in patients with COPD Participating in pulmonary rehabilitation. *Copd J Chronic Obstr Pulm Dis*. 2014;11(4):381-387. doi:http://dx.doi.org/10.3109/15412555.2013.836171.

75. Marin JM, Carrizo SJ, Casanova C, et al. Prediction of risk of COPD exacerbations by the BODE index. *Respir Med*. 2009;103(3):373-378. doi:http://dx.doi.org/10.1016/j.rmed.2008.10.004.

76. Ko FW, Tam W, Tung AH, et al. A longitudinal study of serial BODE indices in predicting mortality and readmissions for COPD. *Respir Med*. 2011;105(2):266-273. doi:http://dx.doi.org/10.1016/j.rmed.2010.06.022.

77. Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. *Am J Med Sci*. 2010;339(1):10-14. doi:http://dx.doi.org/10.1097/MAJ.0b013e3181bb8111.

78. Brusse-Keizer M, van der Palen J, van der Valk P, Hendrix R, Kerstjens H. Clinical predictors of exacerbation frequency in chronic obstructive pulmonary disease. *Clin Respir J*. 2011;5(4):227-234. doi:10.1111/j.1752-699X.2010.00234.x.

79. Takahashi T, Muro S, Tanabe N, et al. Relationship between periodontitis-related antibody and frequent exacerbations in chronic obstructive pulmonary disease. *PLoS One*. 2012;7(7):e40570. doi:10.1371/journal.pone.0040570.

80. Azarisman MS, Fauzi MA, Faizal MP, Azami Z, Roslina AM, Roslan H. The SAFE (SGRQ score, air-flow limitation and exercise tolerance) Index: a new composite score for the stratification of severity in chronic obstructive pulmonary disease. *Postgrad Med J*. 2007;83(981):492-497.

81. Motegi T, Jones RC, Ishii T, et al. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int J Copd*. 2013;8:259-271. doi:http://dx.doi.org/10.2147/COPD.S42769.

82. Gudmundsson G, Gislason T, Janson C, et al. Risk factors for rehospitalisation in COPD: role of health status, anxiety and depression. *Eur Respir J*. 2005;26(3):414-419. doi:10.1183/09031936.05.00078504.

83. Niewoehner DE, Lokhnygina Y, Rice K, et al. Risk indexes for exacerbations and hospitalizations due to COPD. *Chest*. 2007;131(1):20-28. doi:10.1378/chest.06-1316.

84. Jacob A, Laurin C, Boudreau M, Lavoie KL, Bacon SL. The impact of body mass index on chronic obstructive pulmonary disease exacerbations. *Can Respir J*. 2013;18(2):237-242.

85. Suetomo M, Kawayama T, Kinoshita T, et al. COPD assessment tests scores are associated with exacerbated chronic obstructive pulmonary disease in Japanese patients. *Respir Investig*. 2014;52:288-295.

86. Parshall MB, Mapel DW, Rice L, Williams A, O'Reilly J. Predictive validity of short-form health survey 36 items scales for chronic obstructive pulmonary disease exacerbation. *Hear Lung J Acute Crit Care*. 2008;37(5):356-365.

87. Bertens LC, Reitsma JB, Moons KGM, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Copd*. 2013;8:493-499. doi:http://dx.doi.org/10.2147/COPD.S49609.

88. Garcia-Aymerich J, Farrero E, Félez M a, Izquierdo J, Marrades RM, Antó JM. Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. *Thorax*. 2003;58(2):100-105. doi:10.1136/thorax.58.2.100.

89. Chen Y. Readmission in Taiwan. *West J Nurs Res*. 2006;28(1):105-124.

90. Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2014;11(2):149-157. doi:http://dx.doi.org/10.1513/AnnalsATS.201307-243OC.

91. Omachi TA, Yelin EH, Katz PP, Blanc PD, Eisner MD. The COPD severity score: a dynamic prediction tool for health-care utilization. *Copd J Chronic Obstr Pulm Dis*. 2008;5(6):339-346. doi:http://dx.doi.org/10.1080/15412550802522700.

92. Lee SD, Huang MS, Kang J, et al. The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. *Respir Med*. 2014;108(4):600-608.

93. Amalakuhan B, Kiljanek L, Parvathaneni A, Hester M, Cheriyath P, Fischman D. A prediction model

for COPD readmissions: catching up, catching our breath, and improving a national problem. *J Community Hosp Intern Med Perspect*. 2012;2(1):9915. doi:http://dx.doi.org/10.3402/jchimp.v2i1.9915.

94. Ong KC, Earnest A, Lu SJ. A multidimensional grading system (BODE index) as predictor of hospitalization for COPD. *Chest*. 2005;128(6):3810-3816.

95. Echave-Sustaeta J, Comeche Casanova L, Garcia Lujan R, Sayas Catalan J, Gomez de la Camara A, Lopez Encuentra A. Prognosis following acute exacerbation of COPD treated with non-invasive mechanical ventilation. *Arch Bronconeumol*. 2010;46(8):405-410. doi:http://dx.doi.org/10.1016/j.arbres.2010.03.004.

96. Decramer M. Management of COPD according to guidelines. A national survey among Belgian physicians. *Monaldi Arch Chest Dis*. 2003.

97. Rutschmann OT. Knowledge of guidelines for the management of COPD: a survey of primary care physicians. *Respir Med*. 2004.

98. Puhan MA, Zoller M, Ter Riet G. COPD : more than respiratory (comment). *Lancet (comment)*. 2008;371:26-27.

99. Fritsch K, Jacot M-L, Klarer A, et al. Adherence to the Swiss guidelines for management of COPD: experience of a Swiss teaching hospital. *Swiss Med Wkly*. 2005;135(7-8):116-121. doi:2005/07/smw-10844.

100. Calverley PMA, Anderson JA, Celli B, et al. Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2007;356(8):775-789.

101. Vestbo J, Anderson JA, Brook RD, et al. Fluticasone furoate and vilanterol and survival in chronic obstructive pulmonary disease with heightened cardiovascular risk (SUMMIT): A double-blind randomised controlled trial. *Lancet*. 2016;387(10030):1817-1826. doi:10.1016/S0140-6736(16)30069-1.

102. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet*. 2007;370(9589):797-799. doi:10.1016/S0140-6736(07)61383-X.

103. Fabbri LM, Luppi F, Beghé B, Rabe KF. Complex chronic comorbidities of COPD. *Eur Respir J*. 2008;31(1):204-212. doi:10.1183/09031936.00114307.

104. Hamburg MA, Collins FS. The Path to Personalized Medicine. *N Engl J Med*. 2010;363(4):301-304. doi:10.1056/NEJMp1002530.

105. Weston AD, Hood L. Systems Biology , Proteomics , and the Future of Health Care : Toward Predictive , Preventative , and Personalized Medicine Introduction : Paradigm Changes in Health Care. *J Proteome Res*. 2004;3(2):179-196. doi:10.1021/pr0499693.

106. Agusti A. The path to personalised medicine in COPD. *Thorax*. 2014;0(4):1-8. doi:10.1056/NEJMp1002530.

107. Agusti A, Sobradillo P, Celli BR. Addressing the complexity of chronic obstructive pulmonary disease: from phenotypes and biomarkers to scale-free networks, systems biology, and P4 medicine. *Am J Respir Crit Care Med*. 2011;183(9):1129-1137. doi:10.1164/rccm.201009-1414PP.

108. Decramer M, Vestbo J. *Global Initiative for Chronic Obstructive Lung Disease (Updated 2014)*.; 2014.

109. Puhan MA, Bachmann LM, Kleijnen J, Ter Riet G, Kessels AG. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Med*. 2009;7:2. doi:10.1186/1741-7015-7-2.

110. Hingorani ADA, Windt DVD a Van Der, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj*. 2013;346(February 2013):e5793. doi:10.1136/bmj.e5793.

111. Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. *BMC Med*. 2013;11(1):7. doi:10.1186/1741-7015-11-7.

112. *Systematic Reviews (CRD Guidance)*. Center for Reviews and Dissemination, University of York; 2009.

113. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *Annu Intern Med*. 2009;151(4):264-269. doi:10.1371/journal.pmed1000097.

114. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100. doi:10.1371/journal.pmed.1000100.

115. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55-64. doi:10.7326/M14-0697.

116. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

117. Cochrane Statistical Methods group and Cochrane Bias Methods Group. *Assessing Risk of Bias in Included Studies*.; 2011.

118. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-781.

119. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690. doi:10.1136/heartjnl-2011-301246.

120. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research 4: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(7709):1487-1490. doi:10.1136/bmj.b606.

121. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61(11):1085-1094. doi:10.1016/j.jclinepi.2008.04.008.

122. Alonzo TA. Clinical Prediction Models: A Practical Approach to Development, Validation , and Updating (Book Review). *AJE (American J Epidemiol*. 2009;170(4):90033. doi:10.1093/aje/kwp129.

123. Khalili D, Hadaegh F, Soori H, Steyerberg EW, Bozorgmanesh M, Azizi F. Clinical Usefulness of the Framingham Cardiovascular Risk Profile Beyond Its Statistical Performance The Tehran Lipid and Glucose Study. *AJE (American J Epidemiol*. 2012;176(3):177-186. doi:10.1093/aje/kws204.

124. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Risk Factor Models for Neurodevelopmental Outcomes in Children Born Very Preterm or With Very Low Birth Weight : A Systematic Review of Methodology and Reporting. *AJE (American J Epidemiol*. 2017;185(7):601-612. doi:10.1093/aje/kww135.

125. Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF. Predicting Risk of Type 2 Diabetes Mellitus with Genetic Risk Models on the Basis of Established Genome-wide Association Markers : A Systematic Review. *AJE (American J Epidemiol*. 2013;178(8):1197-1207. doi:10.1093/aje/kwt123.

126. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev*. 2017;26(143):1-13. doi:10.1183/16000617.0061-2016.

127. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *AJE (American J Epidemiol*. 2006;163(9):783-789. doi:10.1093/aje/kwj093.

128. Collins GS, Moons KGM. Comparing risk prediction models. *Br Med J*. 2012;344(may24 2):e3186-e3186. doi:10.1136/bmj.e3186.

129. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733-2749. doi:10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.

130. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(November 2001):2313-2324. doi:10.1002/sim.1201.

131. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105-3124. doi:10.1002/sim.1875.

132. Lu G, Ades a. E. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *J Am Stat Assoc*. 2006;101(474):447-459. doi:10.1198/016214505000001302.

133. Salanti G, Higgins J, Ades a E, Ioannidis JP a. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008;17(3):279-301. doi:10.1177/0962280207080643.

134. Lu G, Welton NJ, Higgins J, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Res Synth Methods*. 2011;2(1):43-60. doi:10.1002/jrsm.34.

135. Zarin W, Veroniki AA, Nincic V, et al. Erratum to: Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med*. 2017;15(1):61. doi:10.1186/s12916-017-0832-6.

136. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385.

137. Panagiotou OA. Network Meta-Analysis : Evidence Synthesis With Mixed Treatment Comparison (Book Review). *AJE (American J Epidemiol*. 2015;181(4):288-289. doi:10.1093/aje/kwu471.

138. Takwoingi Y. Empirical Evidence of the Importance of Comparative Studies of Diagnostic Test Accuracy. *Ann Intern Med*. 2013;158.

139. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res*. 2015:0962280215592269-. doi:10.1177/0962280215592269.

140. Manolio TA, Weis BK, Cowie CC, et al. New Models for Large Prospective Studies : Is There a Better Way ? *AJE (American J Epidemiol*. 2012;175(June 2010):859-866. doi:10.1093/aje/kwr453.

141. Franchini a. J, Dias S, Ades a. E, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. *Res Synth Methods*. 2012;3(2):142-160. doi:10.1002/jrsm.1049.

142. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(March):80-97. doi:10.1002/jrsm.1037.

143. Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology*. 1983;148:839-843.

144. Borenstein M, Hedges LV., Higgins J, R. RH. *Introduction to Meta-Analysis*. Wiley; 2011.

145. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2.

146. Nikolakopoulou A, Mavridis D, Salanti G. How to interpret meta-analysis models: fixed effect and random effects meta-analyses. *Evid Based Ment Health*. 2014;17(2).

147. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health*. 2014;17(2):53-57.

148. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*. 2012;12(1):152. doi:10.1186/1471-2288-12-152.

149. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods*. 2013;4(4):291-323. doi:10.1002/jrsm.1085.

150. Cooper NJ, Sutton AJ, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28(15):1982-1998. doi:10.1002/sim.

151. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med*. 2013;11(1):159. doi:10.1186/1741-7015-11-159.

152. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and Technical Challenges in Network Meta-analysis. *Ann Intern Med*. 2013;159:130-137. doi:10.7326/0003-4819-159-2-201307160-00008.

153. Efthimiou O, Debray TP a, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;(November 2014). doi:10.1002/jrsm.1195.

154. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins J. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014;9(7):e99682. doi:10.1371/journal.pone.0099682.

155. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971-980. doi:10.1093/aje/kwq223.

156. Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-1573. doi:10.1002/sim.1187.

157. Dias S, Welton NJ, Caldwell DM, Ades a. E. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(November 2008):932-944. doi:10.1002/sim.3767.

158. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley Serie in Probability and Mathematical Statistics; 1987. doi:10.1002/9780470316696.

159. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol*. 2010;10(1):112. doi:10.1186/1471-2288-10-112.

160. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for Misclassified Outcomes in Binary Regression Models Using Multiple Imputation With Internal Validation Data. *AJE (American J Epidemiol*. 2013;177(9):904-912. doi:10.1093/aje/kws340.

161. Groenwold RHH, Donders ART, Roes KCB, Harrell FE, Moons KGM. Dealing With Missing Outcome Data in Randomized Trials and Observational Studies (Practice of Epidemiology). *AJE (American J Epidemiol*. 2012;175(3):210-217. doi:10.1093/aje/kwr302.

162. White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985. doi:The Stata Journal.

163. Pepe MS. Problems With Risk Reclassification Methods for Evaluating Prediction Models. *AJE (American J Epidemiol*. 2011;173(11):1327-1335. doi:10.1093/aje/kwr013.

164. Harrell FE. *Regression Modelling Strategies*. (Bickel P, Diggle P, Feinberg SE, Gather U, Olkin I, Zeger S, eds.). Springer - Statistics for Biology and Health; ISBN: 978-3-319-19424-0; 2015.

165. Laupacis A, Wells G, Richardson W. S, Tugwell P. Users' guides to the medical literature: V. How to use an article about prognosis. *Jama*. 1994;272(3):234-237. doi:doi:10.1001/jama.1994.03520030076032.

166. Puhan MA, Yu T, Stegeman I, Varadhan R, Singh S, Boyd CM. Benefit-harm analysis and charts for individualized and preference-sensitive prevention: example of low dose aspirin for primary prevention of cardiovascular disease and cancer. *BMC Med*. 2015;13(250):11. doi:10.1186/s12916-015-0493-2.

167. Hayes DF, Markus HS, Leslie RD, Topol EJ. Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC Med*. 2014;12(37):8. doi:10.1186/1741-7015-12-37.

168. Nickel CH, Bingisser R, Morgenthaler NG. The role of copeptin as a diagnostic and prognostic biomarker for risk stratification in the emergency department. *BMC Med*. 2012;10(7). doi:10.1186/1741-7015-10-7.

169. Vuong K, McGeechan K, Armstrong BK, Cust AE. Risk Prediction Models for Incident Primary Cutaneous Melanoma: A Systematic Review. *JAMA Dermatology*. 2014;150(4):434. doi:10.1001/jamadermatol.2013.8890.

170. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA - J Am Med Assoc*. 2011;306(15):1688-1698.

171. Yourman LC, Lee SJ, Schonberg M a, Widera EW, Smith AK. Prognostic Indices for Older Adults: A Systematic Review. *JAMA*. 2012;307(2):182-192.

172. Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9(103):14. doi:10.1186/1741-7015-9-103.

173. Esteban C, Quintana JM, Moraza J, et al. BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice? *BMC Med*. 2010;8:28. doi:10.1186/1741-7015-8-28.

174. Quintana JM, Esteban C, Unzurrunzaga A, et al. Predictive score for mortality in patients with COPD exacerbations attending hospital emergency departments. *BMC Med*. 2014;12:66. doi:http://dx.doi.org/10.1186/1741-7015-12-66.

175. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research 2: Developing a prognostic model. *BMJ (Research Methods Reporting)*. 2009;338(june):1373-1377. doi:10.1136/bmj.b604.

176. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med*. 2010;8(21):11. doi:1741-7015-8-21 [pii] 10.1186/1741-7015-8-21.

177. Hemingway H. Ten steps towards improving prognosis research Problems with prognosis research. *BMJ*. 2014;4184(December 2009):1-10.

178. Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med*. 2016;14(104):15. doi:10.1186/s12916-016-0650-2.

179. Thangaratinam S, Allotey J, Marlin N, et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med*. 2017;15(68):11. doi:10.1186/s12916-017-0827-3.

180. Haile SR, Guerra B, Soriano JB, Puhan MA. Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores. *Stat Methods Med Res*. 2017;17(172):1-12. doi:(submitted).

181. Kessels AG, Riet G, Puhan MA, Kleijnen J, Bachmann LM, Minder C. A simple regression model for network meta-analysis. *OA Epidemiol*. 2013;1(1):1-8.

182. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med*. 2011;9(79):5. doi:10.1186/1741-7015-9-79.

183. Salanti G, Marinho V, Higgins J. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol*. 2009;62(8):857-864. doi:10.1016/j.jclinepi.2008.10.001.

184. Sauter R, Held L. Network meta-analysis with integrated nested Laplace approximations. *Biometrical J*. 2015;57(6):1038-1050. doi:10.1002/bimj.201400163.

185. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating (ROC) Curvel Characteristic. *Radiology*. 1982;143:29-36.

186. Caldwell DM, Ades a E, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj*. 2005;331(7521):897-900. doi:10.1136/bmj.331.7521.897.

187. Mills EJ, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. *Jama*. 2012;308(12):1246-1253. doi:10.1001/2012.jama.11228.

188. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *Bmj*. 2017;356(i6460):11. doi:10.1136/bmj.i6460.

189. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: An overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-171. doi:10.1016/j.jclinepi.2010.03.016.

190. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15(58):9. doi:10.1186/s12874-015-0060-8.

191. Ransohoff DF, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-930.

192. Ioannidis JP a, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. *JAMA - J Am Med Assoc*. 2010;303(16):1646-1647. doi:10.1001/jama.2010.503.

193. Divo M, Cote C, De Torres JP, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155-161. doi:10.1164/rccm.201201-0034OC.

194. Reilly BM, Evans AT, Schaider JJ, et al. Impact of a Clinical Decision Rule in the Emergency Department. *JAMA - J Am Med Assoc*. 2002;288(3):342-350.

195. McGinn TG, Guyatt GH, Wyer PC, David Naylor C, Stiell IG, Scott Richardson W. Users' Guides to the Medical Literature. *J Am Med Assoc*. 2000;284(1):79-84. doi:10.1001/jama.284.1.79.

196. Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA - J Am Med Assoc*. 2009;302(21):2345-2352. doi:10.1001/jama.2009.1757.

197. Wedzicha JA, Brill SE, Allinson JP, Donaldson GC. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. *BMC Med*. 2013;11(181):10. doi:10.1186/1741-7015-11-181.

198. Parkes G. Using risk scores with patients. *Bmj*. 2010;340(jun22 3):c3327-c3327. doi:10.1136/bmj.c3327.

199. Steyerberg EW, Bleeker S., Moll H a, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441-447. doi:10.1016/S0895-4356(03)00047-7.

200. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005.

201. Reilly BM, Evans AT. Academia and clinic. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.

202. Held U, Garcia-Aymerich J, Ter Riet G, Moons KGM, Puhan MA. Missing variables in risk prediction models. *AJE (American J Epidemiol*. 2012. doi:10.1002/elan.

203. Harrell FE. *Regression Modeling Strategies*. Springer; 2015.

204. Snell KIE, Hua H, Debray TPA, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2016;69:40-50. doi:10.1016/j.jclinepi.2015.05.009.

205. Blattenberger G, Lad F. Separating the Brier Score into Calibration and Refinement Components : A Graphical Exposition Author (s): Gail Blattenberger and Frank Lad Source : The American Statistician , Vol . 39 , No . 1 (Feb ., 1985), pp . 26-32 Published by : Taylor & Franci. *Am Stat*. 1985;39(1):26-32.

206. Murphy AH. A New Vector Partition of the Probability Score. *J Appl Metereology*. 1973;12:595-600.

CHAPTER 1: General Introduction

CHAPTER 2: Systematic Review of Prediction Models

Prediction models for exacerbations in patients with COPD: A systematic review

Beniamino Guerra[1], Violeta Gaveikaite[1], Camilla Bianchi[1], Milo A. Puhan[1]

Affiliations: [1]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Eur Respir Rev 2017; 26: 160061

CHAPTER 2: Systematic Review of Prediction Models

Abstract

Personalised medicine aims to tailor medical decisions to the individual patient. A possible approach is to stratify patients according to the risk of adverse outcomes such as exacerbations in COPD. Risk-stratified approaches are particularly attractive for drugs like inhaled corticosteroids or phosphodiesterase-4 inhibitors that reduce exacerbations but are associated with harms. However, it is currently not clear which models are best to predict exacerbations in patients with COPD. Therefore, our aim was to identify and critically appraise studies on models that predict exacerbations in COPD patients. Out of 1382 studies, 25 studies with 27 prediction models were included. The prediction models showed great heterogeneity in terms of number and type of predictors, time horizon, statistical methods and measures of prediction model performance. Only 2 out of 25 studies validated the developed model, and only 1 out of 27 models provides estimates of individual exacerbation risk, only 3 out of 27 prediction models used high quality statistical approaches for model development and evaluation. Overall, none of the existing models fulfilled the requirements for risk-stratified treatment to personalise COPD care. A more harmonised approach to develop and validate high quality prediction models is needed to move personalised COPD medicine forward.

This article has supplementary material available

Support statement: The systematic review was supported by a grant of the University of Zurich

Conflict of interest: None declared

Introduction

Personalised medicine aims to tailor medical decisions to the individual patient [104,105]. The interest in personalised respiratory medicine has risen recently but it has not been introduced much into practice yet [106,107]. For patients with Chronic Obstructive Pulmonary Disease (COPD) [108], a possible approach for personalising medical treatments is to stratify patients according to the risk of exacerbations in order to prescribe treatments such as inhaled corticosteroids or phosphodiesterase-4 inhibitors only if their benefits in terms of reduced risk of exacerbations [109] are expected to outweigh the harms [17,110,111]. For example, a recent benefit harm assessment of the phosphodiesterase-4 inhibitor roflumilast suggested that the risk for severe exacerbations requiring hospital admissions needs to be at least 20% over one

year so that the expected benefits (in terms of reducing severe exacerbations) overcome the gastrointestinal, psychiatric and neurological side effects of roflumilast [17].

Exacerbations are an ideal target for risk-stratified treatment since it is one of the most important outcomes for COPD patients and avoiding them is likely to lead to a higher health-related quality of life, longer life and less health-care cost. However, a prerequisite for risk-stratified treatment is that the risk of exacerbations can be accurately predicted by a prediction model [1,23,24] that has been thoroughly developed and validated [25–28]. A number of models predicting exacerbations in COPD patients have been published reporting on combinations of information from patient history, clinical characteristics and test results including biomarkers to predict exacerbations. It is still not clear yet, though, which prediction model predicts exacerbations most accurately and is applicable in daily practice. For this reason, along with the lack of other systematic reviews, the aim of this systematic review was to identify and critically appraise studies presenting models predicting exacerbations in COPD patients that may support risk-stratified and personalised treatment.

Materials and Methods

The authors followed the Center for Reviews and Dissemination guidance for the methodology [112] and the PRISMA statement for the reporting [113,114].

Protocol

We wrote a detailed study protocol in advance (see online material). We carefully followed the protocol and recorded any deviations from it.

Search methods

We identified eligible papers through a search of the databases Medline (from 1949), Embase (from 1974) and Scopus (from 1996). The search was performed by an information specialist of the University of Zurich (Zurich, Switzerland). Additional studies were identified through the Pubmed-related articles function and reference list of included studies, author contacts, narrative reviews or the grey literature (reports, dissertation, conference abstracts or papers).

Participants

To be eligible for inclusion, patients were required to have a COPD diagnosis according to GOLD criteria (i.e. the ratio between forced expiratory volume (FEV1) and forced volume capacity (FVC) had to be smaller than 0.7 after bronchodilation).

Outcome definition

The outcome of interest was exacerbation. Exacerbations could be event-based (e.g. course of antibiotics and/or oral corticosteroids or admission to hospital) or symptom-based (patient-reported change in symptoms with or without use of diary charts).

General selection criteria

Publication status, year of publication and language were not subject to exclusion criteria.

Study design

We included studies with a longitudinal design (prospective or retrospective cohorts) or control arms of randomised control trials (that can be regarded as cohort studies). Length of follow-up was not subject to exclusion criteria.

Selection criteria for prediction models

For inclusion, the analysis section of the paper had to refer to a prediction model [1] or multivariable association [115,116] of a set of predictors with the outcome exacerbation. By including also multivariable models without explicit reference to prediction models we broadened the eligibility of models substantially in order to learn as much from the literature as possible. But in order to foresee the use of such multivariable models, that often focused on a single predictor of interest (e.g. a biomarker) while adjusting for other predictors (e.g. previous exacerbations or FEV1% predicted) a requirement for inclusion was that the model also included four commonly used and easily available predictors (i.e. previous exacerbations, smoking, age, FEV1% predicted) beside the predictor of interest. Indeed, analyses not accounting for these four common predictors may over-estimate the predictive value of a particular single predictor (such as a biomarker) and there is general consensus that the use of more sophisticated predictors is justified only if they provide additional value when added to commonly available predictors. A further requirement for the inclusion in the systematic review was the presence of at least one performance of the prediction model (e.g. area under the curve -or AUC- for discrimination). If information needed to decide on inclusion was not available from the papers we contacted the authors up to two times to obtain them. The studies could also be included if exacerbation was not the only outcome of the study.

Procedure

Two review authors (BG and CB) independently assessed titles and abstracts of all references retrieved. Two review authors (BG and VG) independently reviewed full-text versions of potentially relevant studies, and selected the studies. Disagreement was resolved by discussion between the two review authors. If consensus was not reached, a third review author was consulted (MP).

Data extraction and management

Two review authors (BG and VG) independently extracted the following data from included studies: demographic characteristics of the study population, disease severity, clinical settings, definition of the outcome, duration of the follow-up, details of the statistical method as well as of the predictors of the final model. All missing information was searched for in the references indicated in the papers (if available), or asked for by email to the authors. Some missing information was retrieved from pharmaceutical companies involved in the studies (if needed, by formal requests).

Quality assessment concerned 6 categories of potential bias (participant selection as shown in the study flow, measurement of predictors, measurement of outcome (i.e. exacerbation), statistical analysis for model development, performance measures and validation, based on guidance from Cochrane [117], an early version of the PROBAST guidelines (http://www.systematic-reviews.com/probast) and the needs of this particular systematic review. The criteria for rating studies at low, high or unclear risk of bias as well as a description for each bias category of each included study are shown in the online material.

Some studies reported on small variations of the same prediction model. Since these models performed in general very similarly, we considered 1 prediction model per study for the main analysis (for details, see online material). We only considered more than one prediction model per study if they were substantially different, as in Almagro et al. [72], where the predictive performance of the CODEX, ADO and BODEX indices were assessed (thus, we will speak in this systematic review of 25 studies and 27 prediction models).

In order to evaluate the readiness of the prediction models for practice we defined a priori three criteria for the clinical applicability of the models.

1. **Availability predictors**: We deemed the set predictors in each prediction model to be *easily (E)* available if most of them were based on questions or information from medical charts, to be *moderately (M)* available if some (at least 2) were based on tests routinely done in non-specialised and specialised settings and to be *difficult (D)* to be available if at least one of the predictor was based on a test usually performed in specialised settings only (details concerning the assessment are explained later in the text and in the online material).
2. **External validation**: In order to be confidently used in practice prediction models require validations in populations other than the populations in which it was developed. We deemed to have high confidence in the performance if the model had been validated (with a small decrease of performance between derivation and validation cohort) in an external cohort of COPD patients and, accordingly, low confidence if an external validation was lacking.
3. **Practical applicability**: To be useful for risk-stratified treatment in practice we deemed models to be useful if they provided a simple point system like the BODE or ADO indices (e.g. [54,56]) with corresponding risks of exacerbations (e.g. 4 points = 25% probability of exacerbation for a specified time-horizon), an online calculator or other means to easily derive the risk of exacerbations for an individual patient. We deemed prediction models not ready for risk-stratified treatment yet if only the statistical methods (e.g. regression coefficients) were reported.

Statistical analysis

Given the heterogeneity of the studies we deemed meta-analyses not a sensible approach and reported the findings using descriptive summary statistics.

Results

Selection of studies

Figure 1 shows the study selection process and the main reasons for exclusion at the different stages. From the database searches we included 20 from a total of 1345 studies. From additional searches we included another 5 studies and thus a total of 25 papers [68–79,81,82,84–90,92–95] reporting on 27 prediction models (see online material for details concerning each stage of the selection process).

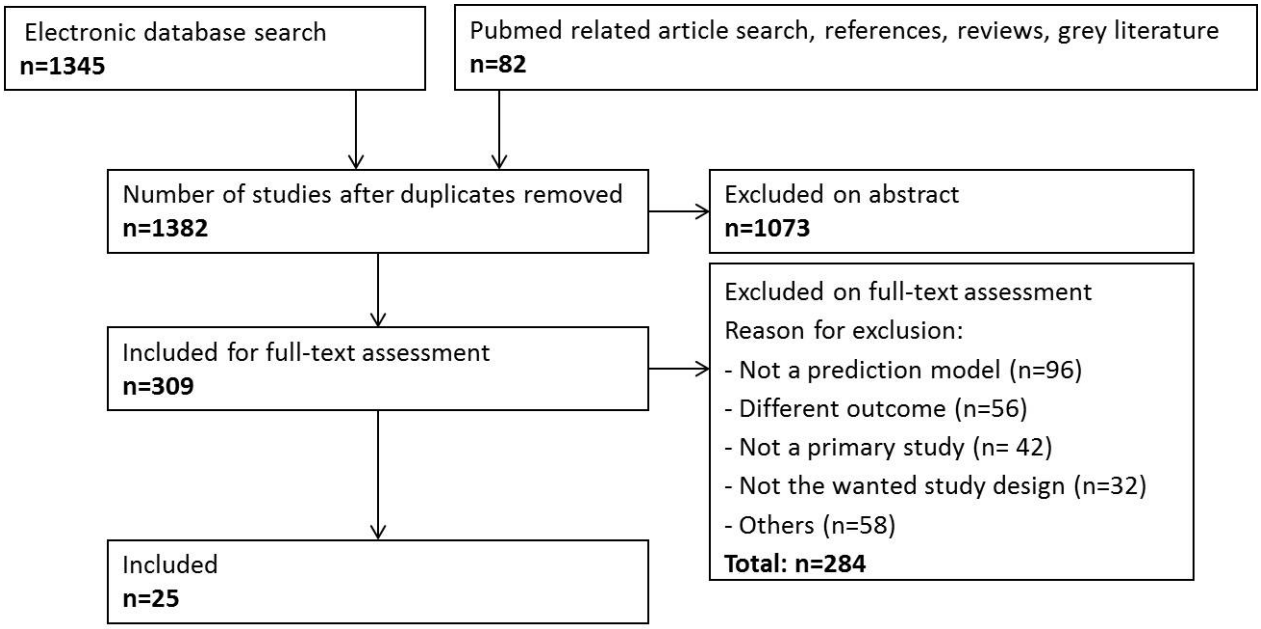


Figure 1 Flow diagram of the study selection process

Study characteristics

The included studies (Table 1) were conducted in countries around the world. They are ordered according to categories of exacerbation incidence. Acknowledging the lack of standard in literature for the individual cut-off value (or more cut-off values) for the frequent exacerbator phenotype, and given the commonly used cut-off of 2 exacerbations per person-years [73,52] we categorised each cohort as low (<1 exacerbation per person-year), moderate (1-2 exacerbations per person-year) or high (>=2 exacerbations per person year) incidence of exacerbations. 3 cohorts with high incidence of exacerbations, 4 cohorts with moderate incidence of exacerbations and 19 cohorts with low incidence of exacerbations were included (for 1 of the 27 cohorts the related data were not retrieved). Sample sizes ranged from 109 to 8020. The definition and measurement of exacerbations was symptom-based in 7 (out of 27 cohorts: 25 derivation cohorts plus 2 validation cohorts), event-based in 17 and unclear definition in 3 cases. Exacerbations were not adjudicated by a committee in any study. The prediction models were mainly based on prospective cohort studies (control arm of a randomised controlled trial for one model [78]), while two prediction models were based on retrospective cohort studies [93,94]). Follow-up periods ranged from 14 days to up to 9 years (the most common follow-up was up to 1 year). 21 out of the 25 included studies had the explicit aim to find a combination of predictors strongly associated with exacerbations, while 4 studies [69,79,82,84] focused on a particular predictor but adjusted for age, FEV1% predicted, smoking and previous exacerbation (making them eligible for the inclusion in this review).

Table 1 Study characteristics

Incidence of exacerbations category	Study (year)	Country	Settings of the care	Cohort study	# patients (included in the cohort study)	# patients (included in the analysis)	Average age	Average FEV1% predicted	Males (%)	Definition of exacerbation	Follow-up (years)	# events in the analysis	Person-years of follow-up	Incidence (exacerbations per year per person)
Low	**Bertens (2013)**	The Netherlands	Outpatient primary care	Prospective	243	240	73	71	68	Event-based	up to 2	>70 [1]	480	>0.15
Low	**Bertens (2013) (validation)**	The Netherlands	Outpatient primary care	Prospective	793	793	67	71	53	Event-based	up to 2	>222 [2]	1586	>0.14
Low	**Parshall (2008)**	UK	Outpatient primary care	Prospective	309	127	70	50	56	Symptom-based [3]	up to 1	>98 [4]	127	>0.77
Low	**Motegi (2013)**	Japan	Outpatient secondary care	Prospective	232	183	71	56	93	Symptom- and event-based	up to 2	193	366	0.53
Low	**Almagro (2014)**	Spain	Outpatient secondary care	Prospective	679	606	73	45	90	Event-based	up to 0.25 and 1	>98 [5]	139.3	>0.70
-	**Almagro (2014) (validation)**	Spain	Outpatient secondary care	Prospective	395	377	72	41	94	Event-based	1	-	-	-
Low	**Jones (2009)**	England	Outpatient secondary care	Prospective	297 [6]	175 [7]	67	42	71	Symptom-based	up to 9	50	1575	0.03
Low	**Suetomo (2014)**	Japan	Outpatient secondary or tertiary care	Prospective	-	123	67	64	87	Symptom-based[8]	up to 1	106 [9]	123	0.87
Low	**Müllerova (2015)**	12 Countries	Secondary and tertiary care	Prospective	2164	2138	63	48	65	Event-based	up to 3	1452	5725 [10]	0.25
Low	**Thomsen (2013)**	Denmark	Outpatient tertiary care	Prospective [11]	8020	6574	67	80	47	Event-based	4 [12]	3083	26296	0.12
Low	**Moberg (2014)**	Denmark	Outpatient tertiary care	Prospective	695	674	69	37	36	Unclear	5.5 [13]	>421 [14]	3822 [15]	>0.11
Low	**Ong (2005)**	Singapore	Outpatient tertiary care	Retrospective	127	127	71	44	91	Symptom-based	1.35 [16]	318 [17]	171 [18]	0.08
Low	**Takahashi (2012)**	Japan	Outpatient tertiary care	Prospective	109	93	73	55	100	Symptom-based	up to 1	92 [19]	93	0.88
Low	**Faganello (2010)**	Brazil	Outpatient tertiary care	Prospective	120	120	65	61	71	Event-based	up to 1	>95 [20]	120	>0.79
Low	**Garcia-Aymerich (2003)**	Spain	Out- and inpatient tertiary care	Prospective	346	312	69	36	92	Symptom-based	1.1 [21]	>197 [22]	343	>0.57
Low	**Ko (2011)**	China	Inpatient tertiary care	Prospective	327	243	74	52	86	Event-based	up to 3	>186 [23]	729	0.26
Low	**Brusse-Keizer (2011)**	The Netherlands	Inpatient tertiary care	Control arm of RCT	121	121	65	58	84	Event-based	up to 1	>62 [24]	121	>0.51
Low	**Echave (2010)**	Spain	Inpatient tertiary care	Prospective	120	93	71	43	89	Event-based	up to 1	>61 [25]	93	>0.65
Low	**Gudmundsson (2005)**	Sweden, Norway, Finland, Iceland, Denmark	Inpatient tertiary care	Prospective	416	406	69	34	49	Event-based	up to 1	>246 [26]	406	>0.61
Low	**Amalakuhan (2012)**	USA	-	Retrospective	-	106	-	-	-	Unclear [27]	up to 1	>100 [28]	106	>0.94
Moderate	**Lee (2014)**	China, Taiwan, Korea, Australia	Outpatient secondary or tertiary care	Prospective	545	495	69	47	88	Symptom-based	up to 0.5	>338 [29]	247.5	>1.37
Moderate	**Moy (2014)**	USA	Outpatient secondary or tertiary care	Prospective	173	167	71	54	99	Event-based [30]	1.25 [31]	263 [30]	210	1.25
Moderate	**Marin (2009)**	Spain	Outpatient tertiary care	Prospective	275	275	65	49	100	Event-based	5.1 [32]	2735 [33]	1402 [34]	1.95
Moderate	**Hurst (2010)**	12 Countries	Inpatient tertiary care	Prospective	2164	2138	63	48	65	Event-based	up to 3	6927 [35]	5725 [36]	1.21
High	**Chen (2005)**	Taiwan	Outpatient secondary care	Prospective	150	143	72	49	73	Event-based	14 days[37]	31	5.5	5.65
High	**Jakob (2013)**	Canada	Outpatient secondary or tertiary care	Prospective	115	115	67	43	47	Event-based	1.5 [38]	683 [39]	207	3.30
High	**Almagro (2006)**	Spain	Inpatient tertiary care	Prospective	156	129	72	36	93	Event-based	1	335 [40]	129	2.60

"-" stays for not reported and not straightforward to evaluate; The incidence of exacerbations category is indicated as low, moderate or high when the exacerbation rate is, respectively, <1 exacerbation per person-year, between 1 and 2 exacerbations per person-year or >2 exacerbations per person-year; FEV1: forced expiratory volume in 1 second; Age and FEV1%predicted are referring to the mean in the study population; [1]70 patients with at least 1 event; [2]222 patients with at least 1 event; [3]Slightly different data are provided for the other outcome analyzed in the paper (health care use); [4]98 patients with at least 1 event; [5]Number of exacerbators to 3 months of follow-up period. Data found in the cited article: CHEST 2012; 142(5):1126–1133; [6]From the reference Am J Respir Crit Care Med Vol 179. pp 369–374, 2009; [7]For which DOSE Index scores were available; [8]The outcome hospital readmission for exacerbation was analyzed as well in the paper; [9]Obtained from the data event per patients in each of the two categories (high-CAT and low-CAT groups); [10]Proxy considering the patients included in the analysis (2138) and the ones completing the three years of follow-up (1679); [11]From a population-based cohort was randomly selected a subgroup of individuals with COPD; [12]Median follow up; [13]Mean follow-up; [14]421 patients with at least 1 event; [15]Obtained multiplying the number of patients in the study by the mean follow-up; [16]Mean follow-up; [17]Calculated considering the mean number of admissions (2.5); [18]Calculated using the mean follow-up value; [19]Proxy obtained using the mean exacerbation frequency per year in the two categories (normal IgG-titer and high IgG-titer); [20]32 patients (27%) experienced 1 episode, 21 patients (18%) 2 episodes, and 7 patients (6%) 3 or more episodes of exacerbation, thus at least 95 exacerbations; [21]Mean follow up; [22]63% of the patients (197) with at least one exacerbation; [23]186 patients with at least 1 readmission for AECOPD; [24]31 patients with at least 2 events; [25]61 patients with at least 1 event; [26]Patients with at least 1 readmission; [27]Presumably event-based; [28]50 patients with multiple (>=2) events; [29]338 patients with at least 1 exacerbation; 226 patients had instead at least 1 moderate to severe exacerbation; [30]Two different outcomes analyzed: number of acute exacerbation and COPD related hospitalization. In this table are only presented the data for the outcome acute exacerbations and not for the outcome COPD related hospitalization; [31]Mean of follow-up; [32]Median follow-up; [33]Obtained multiplying the incidence by person-years; [34]Proxy obtained from the median value for the follow-up; [35]Obtained multiplying the incidence by person-years; [36]Proxy considering the patients included in the analysis (2138) and the ones completing the three years of follow-up (1679); [37]A follow-up of three months was analyzed as well in the paper; [38]Mean follow-up; [39]Obtained multiplying the incidence by person-years; [40]Obtained multiplying the incidence by person-years.

Predictors included in the prediction models

More than 50 different predictors were used across the included prediction models (see online material). Airways obstruction (FEV1% predicted or FEV1 or GOLD stage) was the most common predictor (12 times out of 27 models. The next most common predictors were previous exacerbations (9 times), age (9 times), smoking (8 times) and health-related quality of life (8 times). More than half of predictors were included only once.

Quality assessment

21 out of 25 studies reported with low risk of bias on the study flow and the selection of participants (Figure 2), 23 out of 25 studies were deemed at low risk of bias for how they measured the predictors and 22 out of 25 studies were deemed at low risk of bias for how they measured exacerbations (given our broad definition of exacerbation). 14 studies were at low risk of bias for how the prediction models were developed statistically and how the statistical analysis was performed (the remaining 11 were at high risk). 6 studies were at low risk of bias in terms of the performance measures used (while 19 studies were at high risk). 19 studies out of 25 were of good quality concerning the clinical data (i.e. the three bias categories selection, definition and measurement of predictors and outcomes and in terms of how patients were selected). 3 out of 25 models were of good quality from a statistical point of view (i.e. the two categories statistical method and performance evaluation). Finally, 2 studies [78,93] performed an internal validation [118] and 2 studies [72,87] an external validation (other studies had a validation cohort, but they made a prediction for other outcomes or they did not provide any performance measure for the outcome exacerbation).

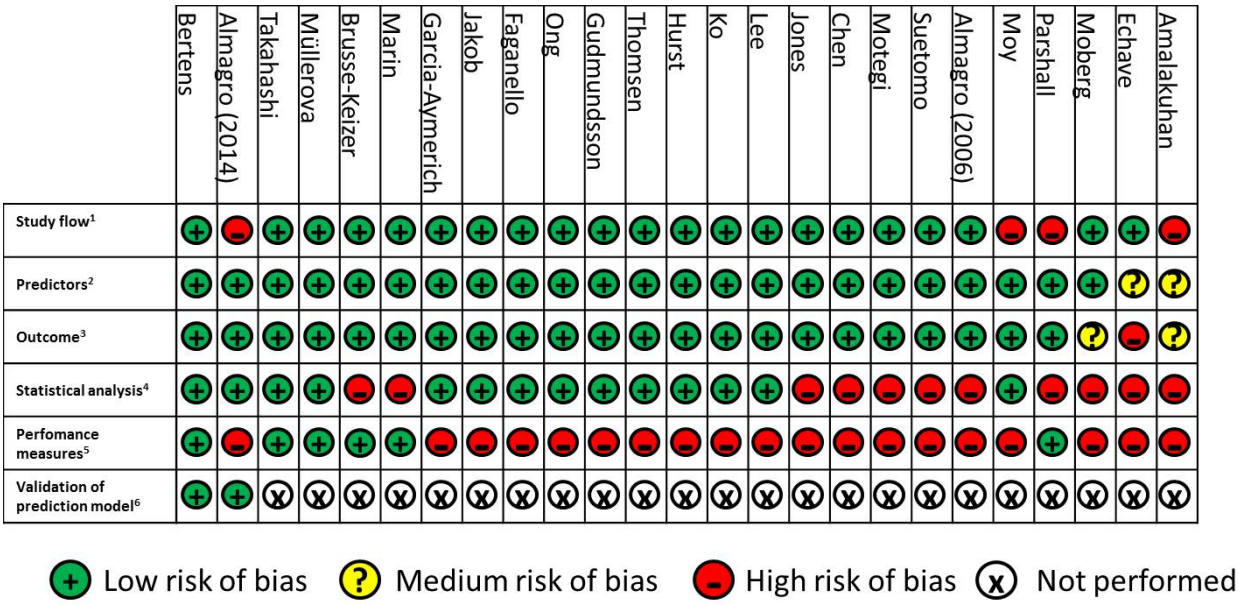


Figure 2 Quality Assessment

1) Study flow description from screening of the patients to patients included in the statistical analysis
2) Definition and measurement of the predictors
3) Definition and measurement of the outcome
4) Statistical method used and eventual procedure of predictors' selection
5) Separated evaluation of discrimination and calibration of the model
6) Validation in an external cohort

Statistical methods

Table 2 shows a description of the 27 prediction models ordered by underlying statistical method (the details of the two validation cohorts are shown as well, for a total of 29 rows); some papers included different analyses, in one case [92] different statistical methods were shown; in order to avoid confusion for the reader, we have included in Table 2 only 1 statistical method per study, apart from the already discussed [72] (where we included the three indices as three independent prediction models). The most common statistical method was logistic regression (11 out of 25 different statistical methods analysed) followed by Cox regression (10), and correlation analysis between an index (or a multivariable regression equation) with the outcome (3). Finally, Poisson regression model, negative binomial regression model, and random forest model were each used once.

Most of the prediction models (18 out of 27) were directly presenting a model with a predefined index or regression equation with predefined predictors. The remaining 9 prediction models used some selection procedure of the variables (i.e. univariable selection process relying on p values, stepwise selection

process relying on p values, combinations of both or selection process driven by the Area Under the Curve).

For 5 prediction models (out of 27) performance related to both discrimination (e.g. AUC) and calibration (e.g. Hosmer-Lemeshow test) were reported (in [87] this is true for both derivation and validation cohort). A measure of discrimination (always AUC) was the most common performance provided (21 times out of 27 prediction models). Measures of overall performance (like R^2 or log-likelihood) and of calibration (Hosmer-Lemeshow p-value or chi-square) were less common (provided, respectively, for 12, 3, 6 and 5 models). The performance provided for the 2 validation cohorts are the same than the ones for their respective derivation cohorts.

Table 2 Description of prediction models ordered by underlying statistical method

Study	Statistical Method	Follow-up (years)	Outcome of the prediction model	Derivation/ validation	Procedure for variable selection	Initial # predictors	Final # predictors	Measure of association	AUC (95%CI)	HL chi-square	HL p value	r^2	-2 Log L
Parshall (2008) [86]	Correlation analysis	up to 1	Exacerbation (symptom-based) [1]	Derivation	Predefined variables [2]	-	1	-	0.65-0.70 [2]	-	0.355-0.974 [2]	0.043-0.081 [2]	-
Marin (2009) [75]	Correlation analysis	5.1 [3]	Outpatient-treated exacerbation [4]	Derivation	Predefined index	-	1	OR	0.78 (0.73-0.84)	-	-	0.31	-
Jones (2009) [70]	Correlation analysis	up to 9	Hospital Admission for exacerbation	Derivation	Predefined index	-	1	-	0.755 [5]	-	-	-	-
Chen (2006) [89]	Logistic regression	14 days [6]	Hospital Readmission for exacerbation	Derivation	Predefined variables	-	9	OR	-	-	-	0.10	139.49
Almagro (2006) [68]	Logistic regression	up to 1	Hospital Readmission for exacerbation	Derivation	Univariable selection	6	3	OR	-	1.9964	0.8496	0.146-0.194 [7]	-
Brusse-Keizer (2011) [78]	Logistic regression	up to 1	Patient with >=2 exacerbations	Derivation [8]	Univariable and stepwise selection	15	2	OR	0.717 (0.595-0.839)	7.512	0.483	0.099-0.15 [9]	-
Faganello (2010) [77]	Logistic regression	up to 1	Exacerbation	Derivation	Predefined index	-	1	OR	0.69 [10]	-	-	-	-
Suetomo (2014) [85]	Logistic regression	up to 1	Exacerbation [11]	Derivation	Predefined variables	3	3	OR	0.77	-	-	-	-
Takahashi (2012) [79]	Logistic regression	up to 1	Patient with >=2 exacerbations	Derivation	Predefined variables	-	11	RR	0.81	66.64	0.56	0.22	-
Hurst (2010) [73]	Logistic regression	up to 1	Increased exacerbation frequency	Derivation	Stepwise selection	31	5	OR	-	-	-	0.22	-
Bertens (2013) (derivation) [87]	Logistic regression	up to 2	Patients with >=1 exacerbation	Derivation	Backward stepwise selection	6	4	OR	0.75 (0.69-0.82)	8.66	0.37	0.26	-
Bertens (2013) (validation)	x [12]	up to 2	Patients with >=1 exacerbation	Validation	x [12]	x [12]	x [12]	x [12]	0.66 (0.62-0.71)	32.98	0.00	0.09	-
Motegi (2013) [81]	Logistic regression	up to 2	Exacerbation	Derivation	P value selection	14	2	OR	0.76-0.78 [13]	-	-	0.15-0.18 [13]	-
Thomsen (2013) [69]	Logistic regression	up to 1 [14]	Frequent exacerbator [15]	Derivation	Predefined variables	-	8-11 [16]	OR	0.71-0.73 [16]	-	-	-	-
Lee (2014) [92]	Logistic regression	up to 0.5	Any exacerbation [17]	Derivation	Predefined variables	-	1-9 [18]	RR	0.64-0.79 [18]	-	-	-	-
Almagro (2014) (derivation) (CODEX) [72]	Cox regression	up to 1 [19]	Hospital Readmission for exacerbation	Derivation	Predefined index	-	1	HR	0.583	-	-	-	-
Almagro (2014) (derivation) (ADO)	Cox regression	up to 1 [19]	Hospital Readmission for exacerbation	Derivation [20]	Predefined index	-	1	HR	0.533	-	-	-	-
Almagro (2014) (derivation) (BODEX)	Cox regression	up to 1 [19]	Hospital Readmission for exacerbation	Derivation [20]	Predefined index	-	1	HR	0.633	-	-	-	-
Almagro (2014) (validation) (CODEX)	x [21]	up to 1	Hospital Readmission for exacerbation	Validation	x [21]	x [21]	x [21]	x [21]	0.590	-	-	-	-
Echave-Sustaeta (2010) [95]	Cox regression	up to 1	Hospital Readmission for exacerbation	Derivation	Univariable and stepwise selection	6	2	HR	0.7601	-	-	0.334	-
Gudmundsson (2005) [82]	Cox regression	up to 1	Hospital Readmission for exacerbation	Derivation	Predefined variables	Unclear	5	HR	-	-	-	-	2175.6
Garcia-Aymerich (2003) [88]	Cox regression	1.1 [22]	Hospital Readmission for exacerbation	Derivation	Univariable selection	23	7	HR	0.71	-	-	-	-
Jakob (2013) [84]	Cox regression	1.5 [23]	Any exacerbation [24]	Derivation	Predefined variables	-	7	RR	-	-	-	-	833.372 [25]
Müllerova (2015) [71]	Cox regression	up to 3	Hospital admission for exacerbation	Derivation	Predefined variables	Unclear	6	HR	0.742 (0.718-0.766) [26]	5.57 [26]	0.696 [26]	0.15 [26]	-
Ko (2011) [76]	Cox regression	up to 3	Hospital Readmission for exacerbation	Derivation	Predefined index	-	1	HR	0.58 [27]	-	-	-	-
Moberg (2014) [74]	Cox regression	5.5 [28]	Hospital admission for exacerbation	Derivation	Univariable selection	Unclear	18	HR	0.62 (0.59-0.65)	-	-	-	-
Ong (2005) [94]	Poisson Regression	1.35 [29]	Hospital Admission for exacerbation	Derivation	Predefined index	-	1	IRR	-	-	-	0.16 [30]	-
Moy (2014) [90]	Negative binomial regression	1.25 [31]	Exacerbation [32]	Derivation	Predefined variables	-	4-5 [33]	RaR	0.59-0.62 [33]	-	-	-	-
Amalakuhan (2012) [93]	Random Forest	up to 1	Hospital Readmission for exacerbation	Derivation [34]	Variables of importance evaluated with AUC	60	5	-	0.75 [35]	-	-	-	-

In this table only the data and analyses related to one outcome (in general the one deemed to be closer to the definition of exacerbation) are reported, with exception of Almagro (2014), where three different rows (for the three different predefined indices: CODEX, ADO, BODEX) are used. "-" stays for not reported and not straightforward to evaluate; AUC: Area Under the Curve; -2 log L: -2 logarithmic likelihood; HL: Hosmer-Lemeshow test; OR: odds ratios; RR: relative risk; RaR: Rate Ratios; HR: hazard ratios; IRR: Incidence Rate Ratios; The number of initial predictors is often a proxy evaluated by the authors of this review; [1]The outcome health care use was analysed as well in the paper; [2]4 domain subscales of the Medical Outcomes Study short-form health survey [36 items] (SF-36) (GH: General Health perceptions; MH: Mental health, RP: Role limitation-Physical, MCS: Correlated Mental-health Score) were evaluated as separated prediction models; [3]Median follow-up; [4]The outcome hospitalisation was analysed as well in the paper; [5] The performance refers to the DOSE index (a performance is provided for MRC Dyspnoea Scale and FEV1%predicted as well in the text); [6]A follow-up of three months was analysed as well; [7]Cox-Snell-Nagelkerke definition of R-square ; [8]To assess over-fitting jackknife cross validation technique was applied to the prediction rule ; [9]Cox-Snell-Nagelkerke definition of R-square; [10]The performance refers to the BODE index (a performance is provided for GOLD stage as well in the text). Other multivariable models are also shown but no performance was provided; [11] The outcome hospital readmission for exacerbation was analysed as well in the paper; [12]Not expected since it refers to validation process; [13]According to the model used (4 different ones are presented); [14]The authors of the paper (Thomsen 2013) report in the online material the data for a median follow-up of 4 year (range of performances with and without the three biomarkers evaluated in the study: 0.92-0.92); [15]At least 2 exacerbations more than 1 year apart; [16]According to the inclusion in the final model of three inflammatory biomarkers as predictors; Net Reclassification Index used to evaluate (40%; 22%-57%) the improvement in performances with the inclusion of the three biomarkers; [17]Other two outcomes (moderate-severe exacerbation with a logistic regression model and time to first exacerbation with Cox regression model) were analysed in the text; [18]According to if CAT was categorised or not and if there was or not adjustment for other variables (age, BMI, duration of COPD, current smoking status, number of comorbidities, history of influenza vaccination, Country); [19]For the derivation cohort the follow-up of 3 months is analysed as well in the paper; [20]Validation is performed but no performance is provided; [21]Not expected since it refers to validation process; [22]Mean follow-up; [23]Mean follow-up; [24]The outcomes outpatients-treated exacerbation and inpatient-treated exacerbation are analysed as well in the paper; [25]Provided as well AIC and SBC, but not reported in the table; [26]Data provided upon request by the authors; [27]The performance refers to the BODE index (a performance is provided for the BODE's 4 component as well in the text); [28]Mean follow-up; [29]Mean follow-up; [30]The performance refers to the BODE index (a performance is provided for the GOLD stage as well in the text); [31]Mean follow-up; [32]The outcome hospital readmission for exacerbation is analysed as well in the paper; [33]Different models evaluated (Step_CRP, STEP_IL-6 or BODE); [34]Internal validation performed; [35]Mean AUC for 200 runs

Clinical applicability of the models

The use of prediction models in practice needs to balance the clinical availability of predictors, i.e. the effort to obtain the information, the easiness with which doctors can obtain a risk for the individual patients and the predictive performance of the models. Ideally, predictors would be easily available, the

model easy to obtain individual probabilities from, and the model would predict the risk of exacerbations accurately as shown by an external validation.

Table 3 shows the assessment of the readiness of the prediction models for clinical practice. The availability of predictors was based on the assessment of the availability of single predictors and how many of them were in different categories of availability (as shown in the online material, "1" refers to a simple test or simple questions or medical charts, "2" refers to routine tests, "3" refers to specialised tests). 12 out of 27 models were deemed to have an easily available set of predictors across non-specialised and specialised health care settings, 4 out of 27 to have an moderately easy available set of predictors and 11 out of 27 to have a set of predictors whose data is difficult to obtain across health care settings. Only 2 models [72,87] can be confidently used in other populations because an external validation was performed to assess the transportability of the prediction model [18]. Also, only 1 study [68] provided a way to easily obtain an estimate of the risk of an exacerbation for an individual patient and thus a basis for risk-stratified treatment. Overall, none of the existing models fulfilled all criteria for being ready for clinical application and use for risk-stratified treatment to personalise COPD care.

Table 3 Readiness of prediction models for clinical practice

Study	Availability predictors	External validation	Practical applicability
Bertens (2013) [87]	E	Y	N
Almagro (2014) [72]	E	Y	N
	E	Y	N
	E	Y	N
Almagro (2006) [68]	D	N	Y[1]
Brusse-Keizer (2011) [78]	E	N[2]	N
Amalakuhan (2012) [93]	E	N[3]	N
Jones (2009) [70]	E	N[4]	N
Echave-Sustaeta (2010) [95]	E	N	N
Jakob (2013) [84]	E	N	N
Motegi (2013) [81]	E	N	N
Lee (2014) [92]	E	N	N
Parshall (2008) [86]	E	N	N
Chen (2006) [89]	M	N	N
Gudmundusson (2005) [82]	M	N	N
Hurst (2010) [73]	M	N	N
Suetomo (2014) [85]	M	N	N
Faganello (2010) [77]	D	N	N
Ko (2011) [76]	D	N	N
Ong (2005) [94]	D	N	N
Garcia-Aymerich (2003) [88]	D	N	N
Marin (2009) [75]	D	N	N
Moberg (2014) [74]	D	N	N
Moy (2014) [90]	D	N	N
Müllerova (2015) [71]	D	N	N
Takahashi (2012) [79]	D	N	N
Thomsen (2013) [69]	D	N	N

The field "Availability predictors" refers to how easy is to obtain the data related to the predictors: E (easy), M (medium), D (difficult); the field "External validation" refers to the reliability of the model in terms of comparison of performance between derivation and eventual validation cohort: Y (yes), N (no). The field "Practical applicability" indicates if it is easy to extract individual likelihoods of exacerbation from the model: Y (yes), N (no). Studies presenting different models are considered only once, with exception of Almagro (2014) where different predefined indices are shown. [1]The paper explains how to you can obtain probabilities starting from the logistic regression equation and indicates how to calculate individual probabilities starting from the predictors; [2]The jackknife cross validation was applied but no performance is indicated; [3]Internal validation was applied but no performance is indicated; [4]Validation is performed for other outcomes.

Discussion

Our systematic review identified 25 studies reporting on 27 statistical prediction models for exacerbation in patients with COPD. The prediction models differ greatly in terms of how they were developed and which predictors and measures for their predictive performance were used. Most studies were of good quality concerning the clinical settings and tests (i.e. selection, definition and measurement of predictors and outcomes and in terms of how patients were selected). However, most of the prediction models were at high risk of bias because unsound statistical methods to develop prediction models, and a lack of validation. The overall assessment of readiness of the 27 prediction models for use in practice showed that none were ready for clinical application.

Strengths and limitations

Strengths of this systematic review are the adherence to rigorous systematic review methodology and reporting guidelines, apart from a thorough search strategy and a great effort for retrieving the needed information from the authors. A limitation could be considered the broad inclusion criteria concerning the definition of exacerbation, potentially introducing heterogeneity among models. Furthermore, the adopted broad definition of prediction model could have allowed the inclusion of studies not meant to concern prediction, but only evaluating the association of an index (or a multivariable regression equation) with the outcome. Nevertheless, we deemed our broader approach suitable in order not to miss prediction models that may be useful for clinical practice. Finally, the big heterogeneity of statistical methods used in literature makes probably not valuable to overall compare all the models even if they are providing the same performance measure (e.g. AUC), since they are often too different in terms of definition of exacerbations, time horizon, statistical method and outcome of the prediction model.

Future research

In order to come up with high-quality prediction models for exacerbations in COPD patients, a standard methodology for developing the models should be adopted [119]. For instance, in certain medical fields some indices were validated and are currently used in clinical setting for risk-stratified prevention and treatment. The cardiovascular field, for example, has a long tradition that started with the Framingham Risk Score predicting the risk of cardiovascular disease [16] and led to clinical guidelines that heavily rely on risk-stratified prevention of cardiovascular disease [7,8]. In COPD, high quality prediction models, for example the BODE and ADO indices, have been developed and externally validated for the outcome of mortality [54–56]. There is also a research need to better understand how prediction models could be made as attractive as possible to use in practice. The optimal balance between availability of predictors, practical applicability and predictive measurement properties is not yet well understood [20,120]. It is paramount that prediction models are validated thoroughly in order to make sure that the risk predictions are accurate across different populations and could be used with confidence for risk-stratified treatment [18,26]. Finally, it would be ideal if the COPD community agreed on a single or very few different exacerbation prediction models since validations and implementation research are more efficient if there is a common prediction model as compared to having many different prediction models [121]. Such a prediction model can always be improved by opportunely updating it (if necessary) in new cohorts [1,29] and by adding promising predictors, but it needs to build upon prior knowledge on other datasets. Of course, separate models are justified if the decisions they inform are distinct, for example, in terms of time horizon.

Conclusions

Overall, none of the existing prediction models fulfilled the criteria for being ready for clinical application and use for risk-stratified treatment to personalise COPD care. The available COPD cohorts contain relevant populations, predictors and exacerbation measurements but a more harmonised approach to develop and validate high quality predictions is needed to move personalised COPD medicine forward.

Acknowledgments

The authors would like to thank Sarah Crook, Laura Barin, Laura Werlen and Alex Marzel (all University of Zurich, Zurich, Switzerland) for their comments.

References

1. Steyerberg EW. *Clinical Prediction Models*. (Gail M, Krickeberg K, Sarnet J, Tsiatis A, Wong W, eds.). Springer - Statistics for Biology and Health - ISBN: 978-1-4419-2648-7; 2010.
2. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: A meta-analysis. *Psychol Assess*. 2000;12(1):19-30. doi:10.1037//1040-3590.12.1.19.
3. Collins GS. The BMJ Gary Collins : Opening up multivariable prediction models. Bmj.
4. Van Gorp MJ, Steyerberg EW, Van Der Graaf Y. Decision Guidelines for Prophylactic Replacement of Bj??rk-Shiley Convexo-Concave Heart Valves: Impact on Clinical Practice. *Circulation*. 2004;109(17):2092-2096. doi:10.1161/01.CIR.0000125853.51637.C8.
5. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation*. 2014;129(25 SUPPL. 1). doi:10.1161/01.cir.0000437738.63853.7a.
6. World Health Organization. *Prevention of Cardiovascular Disease.*; 2007. doi:10.1001/jama.2011.1668.
7. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42(2):517-584. doi:10.1161/STR.0b013e3181fcb238.
8. Expert Panel on Detection, Evaluation and T of HBC in A. Executive summary of the third report of the national cholesterol education program Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Jama*. 2001;285(19):2486-2497. doi:10.1001/jama.285.19.2486.
9. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38. doi:10.1186/1472-6947-6-38.
10. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J Cancer*. 1982;45(3):361-366. doi:10.1038/bjc.1982.62.
11. Di M-Y, Liu H, Yang Z-Y, Bonis PAL, Tang J-L, Lau J. Prediction Models of Mortality in Acute Pancreatitis in Adults. *Ann Intern Med*. 2016;(38). doi:10.7326/M16-0650.
12. Herman WH. Predicting risk for diabetes: Choosing (or building) the right model. *Ann Intern Med*. 2009;150(11):812-814. doi:10.7326/0003-4819-150-11-200906020-00010.
13. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission A Systematic

Review. *JAMA*. 2015;306(15).

14. Ma D, Watson L, Ja R, et al. Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes (Review) Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes. 2011;(11). doi:10.1002/14651858.CD004902.pub4.Copyright.
15. Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. *BMJ*. 2016;354(19):i4338. doi:10.1136/bmj.i4338.
16. Wilson PWF, D'Agostino RB, Levy D, Belanger a. M, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation*. 1998;97(18):1837-1847. doi:10.1161/01.CIR.97.18.1837.
17. Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. *Thorax*. 2014;69(7):616-622. doi:10.1136/thoraxjnl-2013-204155.
18. Justice AC, Covinsky KE, Berlin J a. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-524. doi:199903160-00009 [pii].
19. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353(i2416):11. doi:10.1136/bmj.i2416.
20. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? *Bmj*. 1995;311(7019):1539-1541. doi:10.1136/bmj.311.7019.1539.
21. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007.
22. Collins GS, de Groot J a, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40. doi:10.1186/1471-2288-14-40.
23. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research 1: what , why , and how ? *BMJ*. 2009;338(b375):1317-1320. doi:10.1136/bmj.b375.
24. Steyerberg EW, Moons KGM, Windt DA Van Der, et al. Prognosis Research Strategy (PROGRESS) 3 : Prognostic Model Research. *PLoS Med*. 2013;10(2 (e1001381)):9. doi:10.1371/jour-.
25. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247.
26. Bleeker S., Moll H., Steyerberg EW, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol*. 2003;56(9):826-832. doi:10.1016/S0895-4356(03)00207-5.
27. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research 3: validating a prognostic model. *BMJ*. 2009;338(b605):1432-1435. doi:10.1136/bmj.b605.
28. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.
29. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586. doi:10.1002/sim.1844.
30. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method to update prediction models. *Stat Med*. 2014;Under revi(October).

doi:10.1002/sim.7179.

31. Collins GS, Moons KGM. Comparing risk prediction models Should be routine when deriving a new model for the same purpose. *BMJ*. 2012;3186(May):1-2. doi:10.1136/bmj.e3186.

32. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):93-101. doi:10.1002/bjs.9723.

33. Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM. Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLoS Med*. 2015;12(10):1-12. doi:10.1371/journal.pmed.1001886.

34. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Assessing risk prediction models using individual participant data from multiple studies. *Am J Epidemiol*. 2014;179(5):621-632. doi:10.1093/aje/kwt298.

35. Debray TP a, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med 2013;32*. 2013;32:3158-3180. doi:10.1002/sim.5732.

36. Broeze KA, Opmeer BC, Bachmann LM, et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol*. 2009;9:22. doi:10.1186/1471-2288-9-22.

37. Ahmed I, Debray TP a, Moons KGM, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. *BMC Med Res Methodol*. 2014;14(1):3. doi:10.1186/1471-2288-14-3.

38. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. *Br Med J*. 2015;350(1088):1-9. doi:10.1136/bmj.h1088.

39. Siebeling L, ter Riet G, van der Wal WM, et al. ICE COLD ERIC--International collaborative effort on chronic obstructive lung disease: exacerbation risk index cohorts--study protocol for an international COPD cohort study. *BMC Pulm Med*. 2009;9:15. doi:10.1186/1471-2466-9-15.

40. Soriano JB, Lamprecht B, Ramírez AS, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. *Lancet Respir Med*. 2015;3(6):443-450. doi:10.1016/S2213-2600(15)00157-5.

41. Siebeling L, Puhan MA, Muggensturm P, Zoller M, Ter Riet G. Characteristics of Dutch and Swiss primary care COPD patients - baseline data of the ICE COLD ERIC study. *Clin Epidemiol*. 2011;3:273-283. doi:10.2147/CLEP.S24818.

42. Krumholz HM. Why data sharing should be the expected norm. *Bmj*. 2015;350(feb05 6):h599-h599. doi:10.1136/bmj.h599.

43. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *Bmj*. 2016;353(i3140):11. doi:10.1136/bmj.i3140.

44. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *Bmj*. 2012;344(January):1-10. doi:10.1136/bmj.d7762.

45. Moons KGM, Schuit E. Prediction of cardiovascular disease worldwide. *Lancet Diabetes Endocrinol*. 2015;3(5):309-310. doi:10.1016/S2213-8587(15)00002-9.

46. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework

to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi:10.1016/j.jclinepi.2014.06.018.

47. Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: The COCOMICS study. *Eur Respir J*. 2013;42(2):323-332. doi:10.1183/09031936.00121012.

48. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol*. 1999;28(1):1-9. doi:10.1093/ije/28.1.1.

49. Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. *Heal (San Fr*. 2001;24(2):218-230.

50. Decramer M, Vogelmeier C, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obrstructive Pulmonary Disease (updated 2017). *Glob Initiat Chronic Obstr Lung Dis*. 2011.

51. Celli BR, MacNee W, Agusti A, et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J*. 2004;23(6):932-946. doi:10.1183/09031936.04.00014304.

52. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP.

53. Miravitlles M, Soler-Cataluña JJ, Calle M, et al. Spanish COPD Guidelines (GesEPOC): pharmacological treatment of stable COPD. Spanish Society of Pulmonology and Thoracic Surgery. *Arch Bronconeumol*. 2012;48(7):247-257. doi:10.1016/j.arbres.2012.04.001.

54. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(10):1005-1012. doi:10.1056/NEJMoa021322.

55. Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open*. 2012;2(6):1-10. doi:10.1136/bmjopen-2012-002152.

56. Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet*. 2009;374(9691):704-711. doi:10.1016/S0140-6736(09)61301-5.

57. Soler-Cataluña JJ, Martinez-Garcia MA, Sanchez LS, Tordera MP, Sanchez PR. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. *Respir Med*. 2009;103(5):692-699. doi:http://dx.doi.org/10.1016/j.rmed.2008.12.005.

58. Esteban C, Quintana JM, Aburto M, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. *Respir Med*. 2011;105(11):1662-1670. doi:http://dx.doi.org/10.1016/j.rmed.2011.05.005.

59. Kostianev SS, Hodgev VA, Iluchev DH. Multidimensional system for assessment of COPD patients. Comparison with BODE index. *Folia Medica (Plovdiv)*. 2008;50(4):29-38.

60. Briggs A, Spencer M, Wang H, Mannino D, Sin DD. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. *Arch Intern Med*. 2008;168(1):71-79. doi:http://dx.doi.org/10.1001/archinternmed.2007.37.

61. Schembri S, Anderson W, Morant S, et al. A predictive model of hospitalisation and death from chronic obstructive pulmonary disease. *Respir Med*. 2009;103(10):1461-1467.

doi:http://dx.doi.org/10.1016/j.rmed.2009.04.021.

62. Siebeling L, Musoro JZ, Geskus RB, et al. Prediction of COPD-specific health-related quality of life in primary care COPD patients: a prospective cohort study. *npj Prim Care Respir Med*. 2014;24(14060):7. doi:10.1038/npjpcrm.2014.60.

63. Esteban C, Quintana JM, Aburto M, Moraza J, Capelastegui A. A simple score for assessing stable chronic obstructive pulmonary disease. *QJM*. 2006;99(11):751-759. doi:10.1093/qjmed/hcl110.

64. Boeck L, Soriano JB, Brusse-Keizer M, et al. Prognostic assessment in COPD without lung function: The B-AE-D indices. *Eur Respir J*. 2016;47(6):1635-1644. doi:10.1183/13993003.01485-2015.

65. Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. *Chest*. 2005;127(6):1890-1897. doi:10.1378/chest.127.6.1890.

66. Cardoso F, Tufanin AT, Colucci M, Nascimento O, Jardim JR. Replacement of the 6-min walk test with maximal oxygen consumption in the BODE index applied to patients with COPD: An equivalency study. *Chest*. 2007;132(2):477-482. doi:10.1378/chest.07-0435.

67. Williams JEA, Green RH, Warrington V, Steiner MC, Morgan MDL, Singh SJ. Development of the i-BODE: Validation of the incremental shuttle walking test within the BODE index. *Respir Med*. 2012;106(3):390-396. doi:10.1016/j.rmed.2011.09.005.

68. Almagro P, Barreiro B, Ochoa de Echaguen A, et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. *Respiration*. 2006;73(3):311-317. doi:10.1159/000088092.

69. Thomsen M, Ingebrigtsen TS, Marott JL, et al. Inflammatory biomarkers and exacerbations in chronic obstructive pulmonary disease. *JAMA*. 2013;309(22):2353-2361. doi:http://dx.doi.org/10.1001/jama.2013.5732.

70. Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. *Am J Respir Crit Care Med*. 2009;180(12):1189-1195. doi:http://dx.doi.org/10.1164/rccm.200902-0271OC.

71. Müllerova H, Maselli DJ, Locantore N, et al. Hospitalized Exacerbations of COPD. *CHEST J*. 2015;147(4):999-1007. doi:10.1378/chest.14-0655.

72. Almagro P, Soriano JB, Cabrera FJ, et al. Comorbidities and Short-term Prognosis in Patients Hospitalized for Acute Exacerbation of COPD. *Chest*. 2012;142(5):1126-1133.

73. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2010;363(11):1128-1138. doi:10.1056/NEJMoa0909883.

74. Moberg M, Vestbo J, Martinez G, et al. Validation of the i-BODE index as a predictor of hospitalization and mortality in patients with COPD Participating in pulmonary rehabilitation. *Copd J Chronic Obstr Pulm Dis*. 2014;11(4):381-387. doi:http://dx.doi.org/10.3109/15412555.2013.836171.

75. Marin JM, Carrizo SJ, Casanova C, et al. Prediction of risk of COPD exacerbations by the BODE index. *Respir Med*. 2009;103(3):373-378. doi:http://dx.doi.org/10.1016/j.rmed.2008.10.004.

76. Ko FW, Tam W, Tung AH, et al. A longitudinal study of serial BODE indices in predicting mortality and readmissions for COPD. *Respir Med*. 2011;105(2):266-273. doi:http://dx.doi.org/10.1016/j.rmed.2010.06.022.

77. Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. *Am J Med Sci*. 2010;339(1):10-14. doi:http://dx.doi.org/10.1097/MAJ.0b013e3181bb8111.

78. Brusse-Keizer M, van der Palen J, van der Valk P, Hendrix R, Kerstjens H. Clinical predictors of exacerbation frequency in chronic obstructive pulmonary disease. *Clin Respir J*. 2011;5(4):227-234. doi:10.1111/j.1752-699X.2010.00234.x.

79. Takahashi T, Muro S, Tanabe N, et al. Relationship between periodontitis-related antibody and frequent exacerbations in chronic obstructive pulmonary disease. *PLoS One*. 2012;7(7):e40570. doi:10.1371/journal.pone.0040570.

80. Azarisman MS, Fauzi MA, Faizal MP, Azami Z, Roslina AM, Roslan H. The SAFE (SGRQ score, air-flow limitation and exercise tolerance) Index: a new composite score for the stratification of severity in chronic obstructive pulmonary disease. *Postgrad Med J*. 2007;83(981):492-497.

81. Motegi T, Jones RC, Ishii T, et al. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int J Copd*. 2013;8:259-271. doi:http://dx.doi.org/10.2147/COPD.S42769.

82. Gudmundsson G, Gislason T, Janson C, et al. Risk factors for rehospitalisation in COPD: role of health status, anxiety and depression. *Eur Respir J*. 2005;26(3):414-419. doi:10.1183/09031936.05.00078504.

83. Niewoehner DE, Lokhnygina Y, Rice K, et al. Risk indexes for exacerbations and hospitalizations due to COPD. *Chest*. 2007;131(1):20-28. doi:10.1378/chest.06-1316.

84. Jacob A, Laurin C, Boudreau M, Lavoie KL, Bacon SL. The impact of body mass index on chronic obstructive pulmonary disease exacerbations. *Can Respir J*. 2013;18(2):237-242.

85. Suetomo M, Kawayama T, Kinoshita T, et al. COPD assessment tests scores are associated with exacerbated chronic obstructive pulmonary disease in Japanese patients. *Respir Investig*. 2014;52:288-295.

86. Parshall MB, Mapel DW, Rice L, Williams A, O'Reilly J. Predictive validity of short-form health survey 36 items scales for chronic obstructive pulmonary disease exacerbation. *Hear Lung J Acute Crit Care*. 2008;37(5):356-365.

87. Bertens LC, Reitsma JB, Moons KGM, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Copd*. 2013;8:493-499. doi:http://dx.doi.org/10.2147/COPD.S49609.

88. Garcia-Aymerich J, Farrero E, Félez M a, Izquierdo J, Marrades RM, Antó JM. Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. *Thorax*. 2003;58(2):100-105. doi:10.1136/thorax.58.2.100.

89. Chen Y. Readmission in Taiwan. *West J Nurs Res*. 2006;28(1):105-124.

90. Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2014;11(2):149-157. doi:http://dx.doi.org/10.1513/AnnalsATS.201307-243OC.

91. Omachi TA, Yelin EH, Katz PP, Blanc PD, Eisner MD. The COPD severity score: a dynamic prediction tool for health-care utilization. *Copd J Chronic Obstr Pulm Dis*. 2008;5(6):339-346. doi:http://dx.doi.org/10.1080/15412550802522700.

92. Lee SD, Huang MS, Kang J, et al. The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. *Respir Med*. 2014;108(4):600-608.

93. Amalakuhan B, Kiljanek L, Parvathaneni A, Hester M, Cheriyath P, Fischman D. A prediction model for COPD readmissions: catching up, catching our breath, and improving a national problem. *J Community Hosp Intern Med Perspect*. 2012;2(1):9915.

doi:http://dx.doi.org/10.3402/jchimp.v2i1.9915.

94. Ong KC, Earnest A, Lu SJ. A multidimensional grading system (BODE index) as predictor of hospitalization for COPD. *Chest*. 2005;128(6):3810-3816.

95. Echave-Sustaeta J, Comeche Casanova L, Garcia Lujan R, Sayas Catalan J, Gomez de la Camara A, Lopez Encuentra A. Prognosis following acute exacerbation of COPD treated with non-invasive mechanical ventilation. *Arch Bronconeumol*. 2010;46(8):405-410. doi:http://dx.doi.org/10.1016/j.arbres.2010.03.004.

96. Decramer M. Management of COPD according to guidelines. A national survey among Belgian physicians. *Monaldi Arch Chest Dis*. 2003.

97. Rutschmann OT. Knowledge of guidelines for the management of COPD: a survey of primary care physicians. *Respir Med*. 2004.

98. Puhan MA, Zoller M, Ter Riet G. COPD : more than respiratory (comment). *Lancet (comment)*. 2008;371:26-27.

99. Fritsch K, Jacot M-L, Klarer A, et al. Adherence to the Swiss guidelines for management of COPD: experience of a Swiss teaching hospital. *Swiss Med Wkly*. 2005;135(7-8):116-121. doi:2005/07/smw-10844.

100. Calverley PMA, Anderson JA, Celli B, et al. Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2007;356(8):775-789.

101. Vestbo J, Anderson JA, Brook RD, et al. Fluticasone furoate and vilanterol and survival in chronic obstructive pulmonary disease with heightened cardiovascular risk (SUMMIT): A double-blind randomised controlled trial. *Lancet*. 2016;387(10030):1817-1826. doi:10.1016/S0140-6736(16)30069-1.

102. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet*. 2007;370(9589):797-799. doi:10.1016/S0140-6736(07)61383-X.

103. Fabbri LM, Luppi F, Beghé B, Rabe KF. Complex chronic comorbidities of COPD. *Eur Respir J*. 2008;31(1):204-212. doi:10.1183/09031936.00114307.

104. Hamburg MA, Collins FS. The Path to Personalized Medicine. *N Engl J Med*. 2010;363(4):301-304. doi:10.1056/NEJMp1002530.

105. Weston AD, Hood L. Systems Biology , Proteomics , and the Future of Health Care : Toward Predictive , Preventative , and Personalized Medicine Introduction : Paradigm Changes in Health Care. *J Proteome Res*. 2004;3(2):179-196. doi:10.1021/pr0499693.

106. Agusti A. The path to personalised medicine in COPD. *Thorax*. 2014;0(4):1-8. doi:10.1056/NEJMp1002530.

107. Agusti A, Sobradillo P, Celli BR. Addressing the complexity of chronic obstructive pulmonary disease: from phenotypes and biomarkers to scale-free networks, systems biology, and P4 medicine. *Am J Respir Crit Care Med*. 2011;183(9):1129-1137. doi:10.1164/rccm.201009-1414PP.

108. Decramer M, Vestbo J. *Global Initiative for Chronic Obstructive Lung Disease (Updated 2014)*.; 2014.

109. Puhan MA, Bachmann LM, Kleijnen J, Ter Riet G, Kessels AG. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Med*. 2009;7:2. doi:10.1186/1741-7015-7-2.

110. Hingorani ADA, Windt DVD a Van Der, Riley RD, et al. Prognosis research strategy (PROGRESS) 4:

stratified medicine research. *Bmj*. 2013;346(February 2013):e5793. doi:10.1136/bmj.e5793.

111. Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. *BMC Med*. 2013;11(1):7. doi:10.1186/1741-7015-11-7.

112. *Systematic Reviews (CRD Guidance)*. Center for Reviews and Dissemination, University of York; 2009.

113. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *Annu Intern Med*. 2009;151(4):264-269. doi:10.1371/journal.pmed1000097.

114. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100. doi:10.1371/journal.pmed.1000100.

115. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55-64. doi:10.7326/M14-0697.

116. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

117. Cochrane Statistical Methods group and Cochrane Bias Methods Group. *Assessing Risk of Bias in Included Studies*.; 2011.

118. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-781.

119. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690. doi:10.1136/heartjnl-2011-301246.

120. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research 4: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(7709):1487-1490. doi:10.1136/bmj.b606.

121. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61(11):1085-1094. doi:10.1016/j.jclinepi.2008.04.008.

122. Alonzo TA. Clinical Prediction Models: A Practical Approach to Development, Validation , and Updating (Book Review). *AJE (American J Epidemiol*. 2009;170(4):90033. doi:10.1093/aje/kwp129.

123. Khalili D, Hadaegh F, Soori H, Steyerberg EW, Bozorgmanesh M, Azizi F. Clinical Usefulness of the Framingham Cardiovascular Risk Profile Beyond Its Statistical Performance The Tehran Lipid and Glucose Study. *AJE (American J Epidemiol*. 2012;176(3):177-186. doi:10.1093/aje/kws204.

124. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Risk Factor Models for Neurodevelopmental Outcomes in Children Born Very Preterm or With Very Low Birth Weight : A Systematic Review of Methodology and Reporting. *AJE (American J Epidemiol*. 2017;185(7):601-612. doi:10.1093/aje/kww135.

125. Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF. Predicting Risk of Type 2 Diabetes Mellitus with Genetic Risk Models on the Basis of Established Genome-wide Association Markers :

A Systematic Review. *AJE (American J Epidemiol*. 2013;178(8):1197-1207. doi:10.1093/aje/kwt123.

126. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev*. 2017;26(143):1-13. doi:10.1183/16000617.0061-2016.

127. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *AJE (American J Epidemiol*. 2006;163(9):783-789. doi:10.1093/aje/kwj093.

128. Collins GS, Moons KGM. Comparing risk prediction models. *Br Med J*. 2012;344(may24 2):e3186-e3186. doi:10.1136/bmj.e3186.

129. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733-2749. doi:10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.

130. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(November 2001):2313-2324. doi:10.1002/sim.1201.

131. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105-3124. doi:10.1002/sim.1875.

132. Lu G, Ades a. E. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *J Am Stat Assoc*. 2006;101(474):447-459. doi:10.1198/016214505000001302.

133. Salanti G, Higgins J, Ades a E, Ioannidis JP a. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008;17(3):279-301. doi:10.1177/0962280207080643.

134. Lu G, Welton NJ, Higgins J, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Res Synth Methods*. 2011;2(1):43-60. doi:10.1002/jrsm.34.

135. Zarin W, Veroniki AA, Nincic V, et al. Erratum to: Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med*. 2017;15(1):61. doi:10.1186/s12916-017-0832-6.

136. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385.

137. Panagiotou OA. Network Meta-Analysis : Evidence Synthesis With Mixed Treatment Comparison (Book Review). *AJE (American J Epidemiol*. 2015;181(4):288-289. doi:10.1093/aje/kwu471.

138. Takwoingi Y. Empirical Evidence of the Importance of Comparative Studies of Diagnostic Test Accuracy. *Ann Intern Med*. 2013;158.

139. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res*. 2015:0962280215592269-. doi:10.1177/0962280215592269.

140. Manolio TA, Weis BK, Cowie CC, et al. New Models for Large Prospective Studies : Is There a Better Way ? *AJE (American J Epidemiol*. 2012;175(June 2010):859-866. doi:10.1093/aje/kwr453.

141. Franchini a. J, Dias S, Ades a. E, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. *Res Synth Methods*. 2012;3(2):142-160. doi:10.1002/jrsm.1049.

142. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(March):80-97. doi:10.1002/jrsm.1037.

143. Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology*. 1983;148:839-843.

144. Borenstein M, Hedges LV., Higgins J, R. RH. *Introduction to Meta-Analysis*. Wiley; 2011.

145. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2.

146. Nikolakopoulou A, Mavridis D, Salanti G. How to interpret meta-analysis models: fixed effect and random effects meta-analyses. *Evid Based Ment Health*. 2014;17(2).

147. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health*. 2014;17(2):53-57.

148. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*. 2012;12(1):152. doi:10.1186/1471-2288-12-152.

149. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods*. 2013;4(4):291-323. doi:10.1002/jrsm.1085.

150. Cooper NJ, Sutton AJ, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28(15):1982-1998. doi:10.1002/sim.

151. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med*. 2013;11(1):159. doi:10.1186/1741-7015-11-159.

152. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and Technical Challenges in Network Meta-analysis. *Ann Intern Med*. 2013;159:130-137. doi:10.7326/0003-4819-159-2-201307160-00008.

153. Efthimiou O, Debray TP a, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;(November 2014). doi:10.1002/jrsm.1195.

154. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins J. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014;9(7):e99682. doi:10.1371/journal.pone.0099682.

155. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971-980. doi:10.1093/aje/kwq223.

156. Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-1573. doi:10.1002/sim.1187.

157. Dias S, Welton NJ, Caldwell DM, Ades a. E. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(November 2008):932-944. doi:10.1002/sim.3767.

158. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley Serie in Probability and Mathematical Statistics; 1987. doi:10.1002/9780470316696.

159. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol*. 2010;10(1):112. doi:10.1186/1471-2288-10-112.

160. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for Misclassified Outcomes in

Binary Regression Models Using Multiple Imputation With Internal Validation Data. *AJE (American J Epidemiol*. 2013;177(9):904-912. doi:10.1093/aje/kws340.

161. Groenwold RHH, Donders ART, Roes KCB, Harrell FE, Moons KGM. Dealing With Missing Outcome Data in Randomized Trials and Observational Studies (Practice of Epidemiology). *AJE (American J Epidemiol*. 2012;175(3):210-217. doi:10.1093/aje/kwr302.

162. White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985. doi:The Stata Journal.

163. Pepe MS. Problems With Risk Reclassification Methods for Evaluating Prediction Models. *AJE (American J Epidemiol*. 2011;173(11):1327-1335. doi:10.1093/aje/kwr013.

164. Harrell FE. *Regression Modelling Strategies*. (Bickel P, Diggle P, Feinberg SE, Gather U, Olkin I, Zeger S, eds.). Springer - Statistics for Biology and Health; ISBN: 978-3-319-19424-0; 2015.

165. Laupacis A, Wells G, Richardson W. S, Tugwell P. Users' guides to the medical literature: V. How to use an article about prognosis. *Jama*. 1994;272(3):234-237. doi:doi:10.1001/jama.1994.03520030076032.

166. Puhan MA, Yu T, Stegeman I, Varadhan R, Singh S, Boyd CM. Benefit-harm analysis and charts for individualized and preference-sensitive prevention: example of low dose aspirin for primary prevention of cardiovascular disease and cancer. *BMC Med*. 2015;13(250):11. doi:10.1186/s12916-015-0493-2.

167. Hayes DF, Markus HS, Leslie RD, Topol EJ. Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC Med*. 2014;12(37):8. doi:10.1186/1741-7015-12-37.

168. Nickel CH, Bingisser R, Morgenthaler NG. The role of copeptin as a diagnostic and prognostic biomarker for risk stratification in the emergency department. *BMC Med*. 2012;10(7). doi:10.1186/1741-7015-10-7.

169. Vuong K, McGeechan K, Armstrong BK, Cust AE. Risk Prediction Models for Incident Primary Cutaneous Melanoma: A Systematic Review. *JAMA Dermatology*. 2014;150(4):434. doi:10.1001/jamadermatol.2013.8890.

170. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA - J Am Med Assoc*. 2011;306(15):1688-1698.

171. Yourman LC, Lee SJ, Schonberg M a, Widera EW, Smith AK. Prognostic Indices for Older Adults: A Systematic Review. *JAMA*. 2012;307(2):182-192.

172. Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9(103):14. doi:10.1186/1741-7015-9-103.

173. Esteban C, Quintana JM, Moraza J, et al. BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice? *BMC Med*. 2010;8:28. doi:10.1186/1741-7015-8-28.

174. Quintana JM, Esteban C, Unzurrunzaga A, et al. Predictive score for mortality in patients with COPD exacerbations attending hospital emergency departments. *BMC Med*. 2014;12:66. doi:http://dx.doi.org/10.1186/1741-7015-12-66.

175. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research 2: Developing a prognostic model. *BMJ (Research Methods Reporting)*. 2009;338(june):1373-1377. doi:10.1136/bmj.b604.

176. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med*. 2010;8(21):11. doi:1741-7015-8-21 [pii] 10.1186/1741-

7015-8-21.

177. Hemingway H. Ten steps towards improving prognosis research Problems with prognosis research. *BMJ*. 2014;4184(December 2009):1-10.

178. Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med*. 2016;14(104):15. doi:10.1186/s12916-016-0650-2.

179. Thangaratinam S, Allotey J, Marlin N, et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med*. 2017;15(68):11. doi:10.1186/s12916-017-0827-3.

180. Haile SR, Guerra B, Soriano JB, Puhan MA. Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores. *Stat Methods Med Res*. 2017;17(172):1-12. doi:(submitted).

181. Kessels AG, Riet G, Puhan MA, Kleijnen J, Bachmann LM, Minder C. A simple regression model for network meta-analysis. *OA Epidemiol*. 2013;1(1):1-8.

182. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med*. 2011;9(79):5. doi:10.1186/1741-7015-9-79.

183. Salanti G, Marinho V, Higgins J. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol*. 2009;62(8):857-864. doi:10.1016/j.jclinepi.2008.10.001.

184. Sauter R, Held L. Network meta-analysis with integrated nested Laplace approximations. *Biometrical J*. 2015;57(6):1038-1050. doi:10.1002/bimj.201400163.

185. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating (ROC) Curvel Characteristic. *Radiology*. 1982;143:29-36.

186. Caldwell DM, Ades a E, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj*. 2005;331(7521):897-900. doi:10.1136/bmj.331.7521.897.

187. Mills EJ, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. *Jama*. 2012;308(12):1246-1253. doi:10.1001/2012.jama.11228.

188. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *Bmj*. 2017;356(i6460):11. doi:10.1136/bmj.i6460.

189. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: An overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-171. doi:10.1016/j.jclinepi.2010.03.016.

190. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15(58):9. doi:10.1186/s12874-015-0060-8.

191. Ransohoff DF, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-930.

192. Ioannidis JP a, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. *JAMA - J Am Med Assoc*. 2010;303(16):1646-1647. doi:10.1001/jama.2010.503.

193. Divo M, Cote C, De Torres JP, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155-161. doi:10.1164/rccm.201201-0034OC.

194. Reilly BM, Evans AT, Schaider JJ, et al. Impact of a Clinical Decision Rule in the Emergency Department. *JAMA - J Am Med Assoc*. 2002;288(3):342-350.

195. McGinn TG, Guyatt GH, Wyer PC, David Naylor C, Stiell IG, Scott Richardson W. Users' Guides to the Medical Literature. *J Am Med Assoc*. 2000;284(1):79-84. doi:10.1001/jama.284.1.79.

196. Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA - J Am Med Assoc*. 2009;302(21):2345-2352. doi:10.1001/jama.2009.1757.

197. Wedzicha JA, Brill SE, Allinson JP, Donaldson GC. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. *BMC Med*. 2013;11(181):10. doi:10.1186/1741-7015-11-181.

198. Parkes G. Using risk scores with patients. *Bmj*. 2010;340(jun22 3):c3327-c3327. doi:10.1136/bmj.c3327.

199. Steyerberg EW, Bleeker S., Moll H a, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441-447. doi:10.1016/S0895-4356(03)00047-7.

200. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005.

201. Reilly BM, Evans AT. Academia and clinic. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.

202. Held U, Garcia-Aymerich J, Ter Riet G, Moons KGM, Puhan MA. Missing variables in risk prediction models. *AJE (American J Epidemiol*. 2012. doi:10.1002/elan.

203. Harrell FE. *Regression Modeling Strategies*. Springer; 2015.

204. Snell KIE, Hua H, Debray TPA, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2016;69:40-50. doi:10.1016/j.jclinepi.2015.05.009.

205. Blattenberger G, Lad F. Separating the Brier Score into Calibration and Refinement Components : A Graphical Exposition Author (s): Gail Blattenberger and Frank Lad Source : The American Statistician , Vol . 39 , No . 1 (Feb ., 1985), pp . 26-32 Published by : Taylor & Franci. *Am Stat*. 1985;39(1):26-32.

206. Murphy AH. A New Vector Partition of the Probability Score. *J Appl Metereology*. 1973;12:595-600.

CHAPTER 3: Development of MSC Meta-Analysis

Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores

Sarah R. Haile[1]*, Beniamino Guerra[1]*, Joan B. Soriano[2], Milo A. Puhan[1,3]

* Shared first author

[1]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

[2]Servicio de Neumología, Instituto de Investigación del Hospital Universitario de la Princesa (IISP), Universidad Autónoma de Madrid

[3] Epidemiology & Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA

Haile et al. BMC Medical Research Methodology (2017) 17:172

Abstract

Background: Prediction models and prognostic scores have been increasingly popular in both clinical practice and clinical research settings, for example to aid in risk-based decision making or control for confounding. In many medical fields, a large number of prognostic scores are available, but practitioners may find it difficult to choose between them due to lack of external validation as well as lack of comparisons between them.

Methods: Borrowing methodology from network meta-analysis, we describe an approach to Multiple Score Comparison meta-analysis (MSC) which permits concurrent external validation and comparisons of prognostic scores using individual patient data (IPD) arising from a large-scale international collaboration. We describe the challenges in adapting network meta-analysis to the MSC setting, for instance the need to explicitly include correlations between the scores on a cohort level, and how to deal with many multi-score studies. We propose first using IPD to make cohort-level aggregate discrimination or calibration scores, comparing all to a common comparator. Then, standard network meta-analysis techniques can be applied, taking care to consider correlation structures in cohorts with multiple scores. Transitivity, consistency and heterogeneity are also examined.

Results: We provide a clinical application, comparing prognostic scores for 3-year mortality in patients with chronic obstructive pulmonary disease using data from a large-scale collaborative initiative. We focus on the discriminative properties of the prognostic scores. Our results show clear differences in performance, with ADO and eBODE showing higher discrimination with respect to mortality than other

considered scores. The assumptions of transitivity and local and global consistency were not violated. Heterogeneity was small.
Conclusions: We applied a network meta-analytic methodology to externally validate and concurrently compare the prognostic properties of clinical scores. Our large-scale external validation indicates that the scores with the best discriminative properties to predict 3 year mortality in patients with COPD are ADO and eBODE.

Keywords: prognostic scores, external validation, multiple score comparison, chronic obstructive pulmonary disease

Background

Prediction models, which combine predictors using regression coefficients, and simpler prognostic scores, which typically assign point values to predictors based on prediction models, have become increasingly popular [1,122]. They aid in decision making in public health, clinical research and clinical practice [123] by estimating a person's risk of developing a disease or other outcome. In several medical fields, a variety of prediction models have been developed to assess the individual risk of adverse outcomes. A great example for this was a very recent systematic review regarding validated risk factor models for neurodevelopmental outcomes in children born very preterm or with very low birth weight [124]; 78 original studies (including 222 prediction models) were extracted. Most of the models were not intended to be used for clinical practice and only 4 studies (5%) had performed a validation. Another example regards models predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers; a systematic review deemed to be eligible 21 articles representing 23 studies [125]. Concerning the risk of developing cardiovascular disease, over the past two decades, numerous prediction models have been developed, to estimate the risk of developing cardiovascular disease [19]. Only 36% of them were externally validated and only 19% by independent investigators. In the case of chronic obstructive pulmonary disease (COPD), several prognostic scores have been developed to predict mortality, starting with the BODE score [54]. but scores also exist to predict exacerbations [126], or the course of health-related quality of life [62,70]. Prognostic scores suffer from a reluctance of general practitioners to use them [20,98] as well as from scepticism because they lack internal and external validation which are requirements for generalizability [18,127]. The external validation studies are often simply poorly designed or reported [22]. The lack of comparisons among available prognostic scores provides an additional hurdle to their widespread applicability, as practitioners may not be able to decide among them based on the information available [128].

Luckily, the collection of "big data" [32] and the growing availability of individual patient data (IPD) data analyses [33–38] provide researchers with new opportunities and challenges [43,44]. Furthermore, the call of the medical community for data sharing [42] improves the possibilities of checking a model's predictive performance across clinical settings, populations, and subgroups [43]. The COCOMICS study [47] is a rare example of prognostic scores being directly compared with each other and simultaneously externally

validated after pooling all the databases in a single cohort [128]. Our approach, multiple score comparison network meta-analysis (MSC), extends the simple pooling approach to pool direct comparisons taken from different studies, as a meta-analysis across studies provides in general higher quality information compared to the analysis of a database, constituted pooling together the single studies [48,49]. This methodology allows to take into account heterogeneity of the individual studies and obtain more generalizable results [43].

Methods

Various methodological approaches have been proposed for network meta-analysis for comparison of treatments [129–135], which is sometimes referred to as network meta-analysis, multiple (or mixed) treatment comparisons meta-analysis (MTC meta-analysis) or multiple treatments meta-analysis [136,137]. For diagnostic test performance, the first steps of network meta-analysis were undertaken (e.g. in terms of sensitivity and specificity) [138,139]. No similar methodology exists to compare the performance of prognostic scores or prediction models. Nevertheless, network meta-analysis may provide an attractive solution to the problem of comparing the performance of prognostic scores.

Changing from comparing effects of treatments to comparing performance of prediction models or prognostic scores, however, reveals a number of differences between the two settings, and care must be taken to ensure that the unique features of multiple score comparison (MSC) meta-analysis (as we will refer to this new method) are considered properly in the analysis.

A number of features distinguish a MSC meta-analysis of prognostic scores from a meta-analysis of treatments. In network meta-analysis of treatments outcomes are summarized separately within each treatment arm of a randomized trial, and combined to obtain estimates of treatment effect (for example, mean difference or log odds ratio); instead, the MSC meta-analysis uses performance measures of each score in a cohort that can be calculated on the same sample of patients. Additionally, the number of prognostic scores assessed in a given cohort is not limited by the practicalities of study design, so that it would be easily possible to have more than, say, four scores within one cohort, while such a large number of treatment arms in an RCT is relatively unlikely due to considerations of power and sample size along with practical aspects of conducting clinical trials. Consideration of multi-score studies properly, including the correlations inherent in such comparisons, in MSC is therefore of great importance.

We developed a comprehensive approach to MSC to assess various prediction models using network meta-analysis with individual patient data, providing external validation and concurrent comparison of the scores, and applied it to risk prediction scores for mortality in COPD [40,140]. After careful methodological issues (see also online-only material, where we go deeper into the statistical background) the following approach was developed: we calculated aggregated summary statistics for each cohort and score. Then, we examined the network structure by grouping the cohorts according to which scores could be evaluated. We adapted methodology from network meta-analysis [134] to concurrently externally validate and compare prognostic scores from individual patient data across different cohorts, explicitly including correlations [141] between the scores on a cohort level.

We will also re-interpret NMA as a two-stage meta-regression model, as proposed in [134]:

1. Ordinary meta-analysis to gain the direct estimates for corresponding pooled effect estimates (using the inverse-variance weighted means of the corresponding cohorts). Cohorts at our disposal are classified into "groups" according to which scores it is possible to evaluate by their data.
2. Based on the direct estimates and their variances from the first stage, they obtain to find the optimal estimate of the pooled effect parameters that obeys the fundamental consistency equations. In this stage we merge the group estimates, looking for the weighted least squares solution to the regression problem equation.

The last steps were to confirm that transitivity is a plausible assumption and to check for possible inconsistency and heterogeneity.

Calculation of aggregated summary statistics

First, the performance measure for comparison of the various prognostic scores was defined as the area under the curve (AUC) of the corresponding receiver operating characteristic curve (ROC). This is a graphical plot that illustrates the ability of a binary classifier system (diagnostic or prognostic) as its discrimination threshold is varied (in particular plotting true vs false positive rate). Differences in AUC, denoted ΔAUC, provided an estimate of the relative discrimination ability when comparing scores. For this purpose, we use of a common comparator (CC) model (in our case the GOLD classification, since it is a variable supposed to be present in each COPD cohort); it constitutes a reference value for the performance of other scores, the value from which to subtract the possibly common biases [142].

Variance and covariance estimates for the ΔAUC values were estimated numerically using bootstrapping. We also confirmed consistency of bootstrapped variance estimates to those of the analytical formula for variances of paired differences in AUC [143] (results not shown).

Aggregated data on the cohort level for a cohort with k scores consist therefore of $k - 1$ ΔAUC estimates and a corresponding $(k - 1) \times (k - 1)$ variance-covariance matrix.

To further clarify the methodology, we show the main steps with a small fictional example. Suppose we had 2 cohorts where score A and B could be evaluated (group 1: AB; cohorts P, R), 2 cohorts where A and C could be evaluated (group 2: AC; cohort S, T), 2 cohorts where A, B and D could be evaluated (group 3: ABD; cohorts U, V), and a final 2 cohorts where A, B, C, and D could be evaluated (group 4: ABCD, cohorts X and Y). Let us focus on group 3, constituted by the cohorts X and Y in which the scores A, B, and D can be used. We would obtain performance difference of the scores B and D in comparison to the score A for each of the cohorts, as reported in Table 1.

Table 1 Point estimate of the difference of AUC of the scores B and D with the score A in the group 3 of the fictional example

Cohort	ΔAUC – AB	ΔAUC - AC	ΔAUC - AD
X,	0.05	--	0.07
Y	0.06	--	0.18

Analogously, in group 3, we would obtain variance-covariance matrices, like the ones reported in Table 2.

Table 2 Variance-covariance matrices of the difference of AUC of the scores B and D with the score A in the group 3 of the fictional example.

Cohort X		
	0.0012	0.0005
	0.0005	0.0009

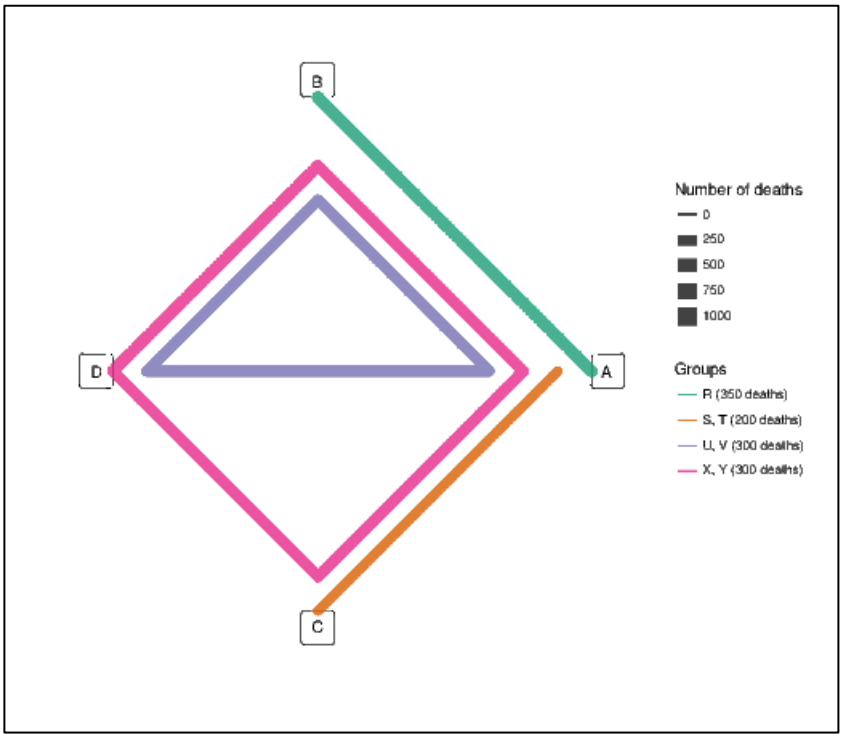
Cohort Y		
	0.0068	0.0051
	0.0051	0.0109

Examination of network structure

Once the aggregated summary statistics were computed, we explored the structure of the network. In a first step, we divided the cohorts into groups based on which sets of scores could be evaluated.

Each group is represented by a polygon, that passes by all the scores (i.e. the vertices) which can be evaluated in the cohorts constituting that group. The thickness of the polygon is directly proportional to the number of deaths in the group (Figure 1).

Figure 1 Network representation of the fictional example



Head-to-head comparisons within a group can be performed between any two scores connected in the same polygon.

For example, in group 4, A and D can be compared because they are both connected by the same polygon, even though there is no line directly connecting the two scores in that group.

According to Table 3, group 1, group 2, group 3 and group 4 have a cumulative number of 4000, 1000, 3000 and 2000 patients, respectively.

Table 3 Group Characteristics of a fictional network (g identifies the group. n is the total number of subjects and d is the total number of deaths in each group. Additional characteristics are also listed: the Q statistic describing heterogeneity has df degrees of freedom, and τ gives the square root of the τ^2 statistics for between cohort heterogeneity)

G	Scores	Cohorts	*n*	*D*	*Q*	*df*	τ
1	A, B	R	4000	350	28	0	0.019
2	A, C	S, T	1000	200	15.5	1	0.014
3	A, B, D	U, V	3000	300	13.4	2	0.004
4	A, B, C, D	X, Y	2000	300	9.6	3	0.036

Multiple score comparison

The method of Lu et al. 2011 [134] was used to perform the multiple score comparison meta-analysis with Der Simonian-Laird random effects [144–147]. This method, which reinterprets frequentist NMA as a two-stage meta-regression model (using inverse variance weighted least squares estimation) , was chosen as, compared to most of the network meta-analytic techniques, it can easily handle multi-score cohorts, and does not make unnecessary simplifications with respect to the correlations inherent in such trials, as discussed above. In the first stage, cohorts in which the same set of scores have been assessed are grouped together and meta-analysed separately.

An estimation T^2 of the between-cohort variance (τ^2) (i.e., the variance of the true performance difference across all studies) is the Der Simonian-Laird method [145] adapted to the network meta-analysis case [134]. Indeed, the *Q* statistic (adapted to network meta-analysis) is referred to a χ distribution with degrees of freedom $df_g=(M_g-1)(N_g-1)$, where M_g is the number of scores compared in the group g and N_g is the number of cohorts belonging to the group *g*. Thus, the degrees of freedom are $df_1=1*0=0$, $df_2=1*1=1$, $df_3=2*1=2$, $df_4=3*1=3$. Table 3 allows us to calculate pooled τ^2 (according to the methods of moments)[144] with which we evaluate the weights used to obtain the weighted average of the performance estimate for the whole network (reported in the first 4 rows in Table 4).

Analogously, extending the definitions from meta-analysis [144] to network meta-analysis [134] we calculate the variables C, Q and τ (τ represents the heterogeneity and deserves further discussion in the text later).

In Stage II the inverse variance weighted least square solution across all groups is found, thus we obtain the performance vector related to each score, best fitting the results of Stage I (for more details see Appendix A). In the last row of Table 4. The final results of the MSC meta-analysis for the fictional example of are reported.

Table 4 Stage I and Stage II Results of the MSC Meta-Analysis in the fictional example of Table 1 (comparison with the A score)

Stage	*G*	B	C	D
I	1	0.09 (0.07, 0.12)		
I	2		0.18 (0.07, 0.29)	
I	3	0.10 (0.08, 0.12)		0.22 (-0.05, 0.50)
I	4	0.08 (0.04, 0..12)	0.15 (0.01, 0.29)	0.18 (0.05, 0.31)
II		0.09 (0.06, 0.13)	0.17 (0.10, 0.25)	0.21 (0.07, 0.35)

Transitivity, heterogeneity and inconsistency. The main assumptions to be met for performing a network meta-analysis are transitivity (a key assumption related to consistency), heterogeneity (differences in estimates of the same treatment or score contrasts coming from different studies) and inconsistency (comparing direct and indirect estimates, sometimes referred to as incoherence) [142,148]. A key assumption of consistency is transitivity (sometimes referred to as similarity [149]) among the treatment effects [133,142,149–153], that is, that indirect comparisons are valid estimates of (unobserved) direct comparisons. Therefore one statistical approach to check for transitivity in our case is to explore the distribution variables giving case-mix across groups [154,155], which we have adopted here using meta-regression. In practice, we used the definition of transitivity from a review paper on the topic [142] better matching our

methodology, namely that the different sets cohorts do not differ with respect to the distribution of variables that could generate case mix variation. Thus, we evaluated by meta-regression analysis [156] the distribution of the variables that could generate case mix variation (like median and variability of age [43], range and variance of obstruction severity (i.e., FEV1% pred.), exercise capacity, size, mortality rate).

In case of variables directly affecting the performance, we used analysis of variance (ANOVA) to see whether the distribution of the identified variables was imbalanced in the groups and could consequently generate imbalance in the performance group by group. In case of homogenous groups, we cannot reject the null hypothesis of transitivity. With this method we assess as well, the eventuality that within-cohort heterogeneity could affect the analysis when "case-mix" is present (i.e. heterogeneity in the variables representing heterogeneity in the cohorts, like FEV%predicted range, that could affect the discriminative properties in the specific cohorts)."

Heterogeneity could be described using a multivariate version of the usual τ^2 statistic, which in the Lu-Ades (2011) approach is considered on a group level at stage 1. They suggest that a pooled τ^2 may be a natural solution to situations where there are singleton groups (i.e. constituting only a cohort).

Inconsistency was primarily assessed visually by comparing direct and indirect comparisons from node-splitting side by side [157]. As a further check of inconsistency, we further examined the Q statistic (that is, the residual sum of squares) which can be used to reject the hypothesis of inconsistency between direct and indirect estimates if Q is greater than the χ^2 statistic with N – K + 1 degrees at freedom at the 100(1 - α)% level [134], where N is the sum of the number of contrasts in each group, and K is total number of scores. Furthermore local consistency could be assessed at the group level by examining residuals and leverage statistics. Furthermore, we considered ways to calculate direct and indirect evidence within the network. Direct comparisons were computed by including only cohorts where both scores under consideration were present, and then performing the usual random effects meta-analysis [144]. However, defining loops of any order for indirect comparisons proved to be difficult in our setting, where the network is highly connected, and most cohorts have between 4 and 9 scores being assessed. Due to the various difficulties presented by studies with multiple scores, we chose to examine inconsistency in the network using "node-splitting" [157]. This approach avoids the need to define loops of any order, and includes all possible indirect evidence.

Consideration of missing data. The main analysis was performed without any imputation technique. A sensitivity analysis, using multiple imputation was also performed and it is shown in the online-only material. The results were not significantly different in the two cases.

COPD Data Description. Following the recommendation for large prospective studies [140], we based our analysis on a large-scale database (provided by the COPD Cohorts Collaborative International Assessment (3CIA) consortium [40]) from a diverse set of 24 cohort studies and 15,762 patients with COPD (1,871 deaths and 42,203 person-years of follow-up). The cohorts were heterogeneous concerning geographic location, sample size, number of events and correspond to a broad spectrum of patients with COPD from primary, secondary and tertiary care settings. Mean FEV1 ranged from 30 to 70% of the predicted values, mean modified Medical Research Council (mMRC) dyspnea scores from 1.0 to 2.8 (the

scale goes from 0 to 4, with 4 being the worst), mean number of exacerbations in the previous year (where available) from 0.2 to 1.7 and mean 6-minute walk distance (where available) from 218 to 487 meters. The follow-up period varied from cohort to cohort, thus we decided to use a minimum common time frame of 3 years. The mean age varies between 58 and 72 years. The outcome of interest was 3-year all-cause mortality. A table summarizing the clinical characteristics of the cohorts is reported in the supplementary material.

Results

To illustrate an MSC meta-analysis, we compared the prognostic ability of various scores to predict mortality in patients with COPD. The COPD Cohorts Collaborative International Assessment (3CIA) [40] initiative contains individual data for around 16000 patients (approx. 70000 person years) with COPD from 26 cohorts in 7 countries. Patients were considered to have COPD if the ratio of forced expiratory volume in one second (FEV1) to forced vital capacity (FVC) was less than 70%, regardless of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) (2007) stage (I-IV) [52]. The minimum required set of variables for each cohort included vital status (up to death, loss to follow-up, or last data collection in June 2013), age, sex, pre-bronchodilator and post-bronchodilator FEV_1 and dyspnoea MRC grade [40]. Most cohorts included many more variables allowing for the calculation of a total of 10 prognostic scores: GOLD (2007), GOLD (2011), updated ADO, BODE, updated BODE, eBODE, BODEx, DOSE, SAFE and optimised B-AE-D [50,54,56,57,64,70,80].

Examination of network structure

We apply the MSC network meta-analysis of prognostic scores for 3-year mortality from the 3CIA data.

Based on the availability of the 10 scores in each cohort, the cohorts could be divided into 6 groups. The network structure is shown in Table 5 and in Figure 2.

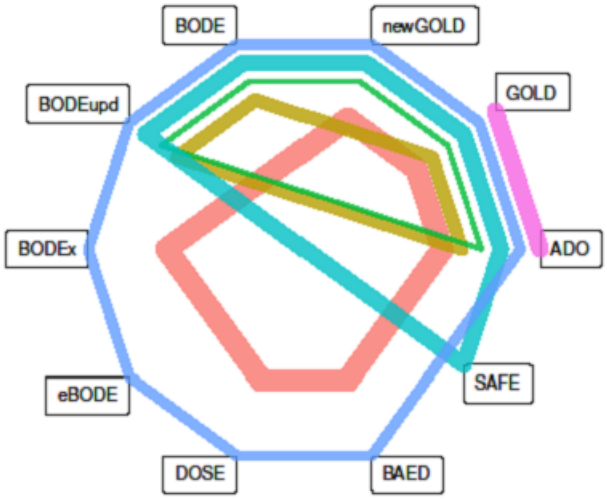


Figure 2 Depiction of network structure with lines weighted by the total number of deaths in the group

Abbreviations: GOLD, Global initiative for chronic Obstructive Lung Disease; BODE, Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODE upd., BODE updated; ADO, Age, Dyspnoea, airflow Obstruction (we use the updated version of the ADO score in our analysis); e-BODE, severe acute exacerbation of COPD plus BODE; BODEx, Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; DOSE, Dyspnoea, Obstruction, Smoking and Exacerbation frequency; SAFE, Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D, Body-mass index, Acute Exacerbations, Dyspnoea.

[a]Cohorts belonging to the ADO or COCOMICS groups are marked with * or † respectively.

[b]The thickness of the lines is proportional to the number of deaths of the respective cohort.

[c]Below we report the composition of the group; each of them is identified by a specific color:

Copenhagen*, HUNT, Japan - SEPOC* (378 deaths in 4323 patients)

Barmelweid*, Basque*, Galdakao†, Pamplona†, Zaragoza I† (215 deaths in 1208 patients)

Mar de Plata Argentina, PACECOPD*, Son Espases Mallorca (61 deaths in 556 patients)

COPDgene (337 deaths in 4484 patients)

Genkols, ICE COLD ERIC, Initiatives BPCO, Sevilla†, Terrassa I†, Terrassa III†, Zaragoza II† (722 deaths in 4346 patients)

La Princesa Madrid, Requena II†, Tenerife†, Terrassa II† (125 deaths in 845 patients)

Table 5 Group Characteristics of the Network

g	Scores	Cohorts	n	#d	Q	df	T
1	GOLD – ADO	Copenhagen*, HUNT, Japan - SEPOC*	4323	378	2.8	3	0
2	GOLD - ADO - BODE – BODEupd	Barmelweid*, Basque*, Galdakao†, Pamplona†, Zaragoza I†	1208	215	15.5	12	0.014
3	GOLD - GOLD (2011) - ADO - BODE - BODEupd	Mar de Plata Argentina, PACECOPD*, Son Espases Mallorca	556	61	10.9	8	0.025
4	GOLD - GOLD (2011) - ADO - BODE - BODEupd - SAFE	COPDgene	4484	337	7.46E-29	0	2.26E-09
5	GOLD - GOLD (2011) - ADO - BODEx - DOSE – BAED	Genkols, ICE COLD ERIC, Initiatives BPCO, Sevilla†, Terrassa I†, Terrassa III†, Zaragoza II†	4346	722	48.1	30	0.011
6	GOLD - GOLD (2011) - ADO - BODE - BODEupd - eBODE - BODEx - DOSE – BAED	La Princesa Madrid, Requena II†, Tenerife†, Terrassa II†	845	125	34.5	24	0.014

Abbreviations: n, number of subjects; d, number of deaths; c, number of cohorts; Q, likelihood ratio statistic; df, degrees of freedom; τ, heterogeneity within the group; GOLD, Global initiative for chronic Obstructive Lung Disease; BODE, Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODE upd., BODE updated; ADO, Age, Dyspnoea, airflow Obstruction (we use in the our analysis the updated version of the ADO score); e-BODE, severe acute exacerbation of COPD plus BODE; BODEx, Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; DOSE, Dyspnoea, Obstruction, Smoking and Exacerbation frequency; SAFE, Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D, Body-mass index, Acute Exacerbations, Dyspnoea.

[a]Cohorts belonging to the ADO or COCOMICS groups are marked with * or † respectively We notice that for group 4 the value of heterogeneity tau is not available (NA); indeed, that is a singleton group, where we cannot evaluate heterogeneity

Even if it would make sense to use absolute performance, we used relative performance of each score in comparison to a Common Comparator score, in order to get rid of possible common biases. We chose as

Common Comparator score the GOLD classification. One cohort (COCOMICS Requena I) was excluded from the analysis because it only had sufficient variables to evaluate a single score (GOLD) and it would not contribute to the analysis. We had to further exclude the cohort A1ATD because there were no cases in the follow-up considered for our analysis (3 years) and the lack of events does not allow calculating an AUC. Of the remaining 24 cohorts, 4 had two scores (GOLD (2007), updated ADO), and the other 20 had between 3 and 9 scores assessed. We note that in no cohort all the 10 scores could be evaluated.

As GOLD (2007) is commonly used to classify the grade of severity of COPD patients, it could be assessed in all cohorts [52]. We note that direct evidence was available for 41 of 45 score comparisons, indirect evidence was available for other 16 comparisons (among which the 4 cases in which the direct comparison was missing).

Multiple score comparison meta-analysis (MSC)

Transitivity, heterogeneity and inconsistency. To examine whether transitivity was fulfilled, we analysed the distribution of a number of possible variables potentially generating case-mix, following epidemiological reasoning and literature (age median and variability [43], FEV1 percent predicted range and variance, mortality percentage, exercise capacity range, number of events) across the groups using meta-regression. For the variables generating case-mix (whose meta-regression analysis were significant), the ANOVA analysis showed that the variables were balanced in the groups. Thus, we cannot reject the null hypothesis of transitivity.

Stage I group level results are presented in the top of Table 6, while the bottom rows show the stage II overall results from the network meta-analysis. GOLD (2007) scores ranged from 0.481 to 0.731, with a median of 0.614, and interquartile range (0.587, 0.641). Of the scores, the one that predicted mortality best was updated ADO with an average AUC 0.083 higher than that of GOLD (2007) (95% confidence interval: 0.069, 0.097), followed by the updated BODE which was associated with a 0.072 better AUC than GOLD (95% confidence interval: 0.051, 0.093) and eBODE (+0.069, 95% confidence interval: 0.044, 0.093). DOSE (+0.027, 95% confidence interval: 0.010, 0.045), optimised B-AE-D (+0.016, 95% confidence interval: -0.007, 0.038) and GOLD (2011) (+0.014, 95% confidence interval: 0.001, 0.028) and showed the worst performance in predicting mortality, only slightly better than GOLD (2007). The other scores, BODE, SAFE and BODEx showed moderate performance, between +0.045 and +0.064 improvement in AUC over GOLD.

Table 6 Stage I and Stage II Results of the MSC Meta-Analysis (Comparison With the GOLD Classification)

Stage	g	ADO	BODEupd	eBODE	BODE	SAFE	BODEx	DOSE	BAED	newGOLD
I	1	0.097 (0.07, 0.123)								
I	2	0.098 (0.057, 0.139)	0.124 (0.078, 0.17)		0.098 (0.059, 0.137)					
I	3	0.044 (-0.03, 0.117)	0.023 (-0.054, 0.099)		0.019 (-0.054, 0.091)					-0.011(-0.053, 0.03)
I	4	0.042 (0.01, 0.074)	0.043 (0.008, 0.078)		0.049 (0.017, 0.081)	0.037 (0.005, 0.069)				-0.008(-0.038, 0.022)
I	5	0.099 (0.076, 0.123)					0.056 (0.035, 0.076)	0.036 (0.015, 0.057)	0.032 (0.005, 0.058)	0.028 (0.008, 0.047)
I	6	0.076 (0.027, 0.126)	0.043 (-0.006, 0.092)	0.048 (0.004, 0.093)	0.043 (0.001, 0.085)		0.030 (-0.015, 0.074)	0.021 (-0.023, 0.065)	-0.017(-0.079, 0.045)	0.008 (-0.031, 0.047)
II		0.083 (0.069, 0.097)	0.072 (0.051, 0.093)	0.069 (0.044, 0.093)	0.064 (0.045, 0.082)	0.052 (0.022, 0.082)	0.045 (0.029, 0.061)	0.027 (0.010, 0.045)	0.016 (-0.007, 0.038)	0.014 (0.001, 0.028)

Abbreviations: MSC, Multiple Score Comparison; GOLD, Global initiative for chronic Obstructive Lung Disease; BODE, Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODE upd., BODE updated; ADO, Age, Dyspnoea, airflow Obstruction (we use in the our analysis the updated version of the ADO score); e-BODE, severe acute exacerbation of COPD plus BODE; BODEx, Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; DOSE, Dyspnoea, Obstruction, Smoking and Exacerbation frequency; SAFE, Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D, Body-mass index, Acute Exacerbations, Dyspnoea.

[a]The first 6 rows show the Stage I results (group by group). The last row shows the Stage II results (namely the final results of the multiple score comparison meta-analysis).

[b]The score ordered by performance of the prognostic scores in the Stage II

Concerning heterogeneity, due to the group containing only a single cohort (group 6), we primarily considered a random effects analysis calculated using a pooled estimate of τ^2 for all groups, which was 0.00015, indicating a relatively low heterogeneity. The results of the MSC network meta-analysis were not substantially different when using the group-specific τ^2 estimates.

Possible inconsistency between direct and indirect comparisons was assessed using the *Q* statistic as described above. Overall, Q for the random effects analysis was 22.1 with 16 degrees of freedom. Keeping in mind that in this case (as in classical network meta-analysis) the inconsistency test has low power, since *Q* was smaller than the corresponding χ^2 statistic of 26.3, we do not reject the hypothesis of consistency ($P = 0.14$).

Both direct and indirect estimates of the score comparisons were calculated using node-splitting [157], and compared visually. The results are similar to each other and to the estimates provided by the network meta-analysis (see supplementary material for further discussion).

Consideration of missing data. As a secondary analysis, the entire meta-analysis was repeated in a multiple imputation framework, as described above. The results were similar to the main analysis without imputation (see supplementary material) [1,158,159].

Discussion

To the best of our knowledge, the MSC meta-analysis proposed in this paper represents the first methodology to evaluate the comparative prognostic properties of prediction models that synthesizes all available (direct and indirect) evidence. The application of the MSC meta-analysis could provide different clinical fields with a clear indication of which is the best-performing prediction model, paving the way for a standardized clinical application. While there are a number of issues when adapting usual NMA methodology to MSC, they can be addressed in a straightforward manner. Multi-score studies are considered in our approach by explicitly using covariance estimates for the various prognostic scores. Calculation of such estimates using bootstrapping may be computationally intensive but is not difficult to implement. The approach presented here can be used to compute prognostic score comparisons for the entire network of evidence, as well as both direct and indirect comparisons between scores.

Despite these adaptations, the results of the MSC meta-analysis are clear, and may be interpreted in a fashion similar to standard network meta-analysis results. The only difference is that the performance measure is not mean difference between treatments, or log odds ratio, but difference in performance measure such as AUC. Measures of heterogeneity and inconsistency can however be calculated and interpreted in the usual fashion [142]. For instance, a definition similar to the one used for the heterogeneity for meta-analysis of direct comparisons, can be used for the heterogeneity of network meta-analysis, adapting a definition used for multi-arm trials to multiple score comparison. Since we have singleton groups in our MSC data (group 6 in our database), it is recommended in our case to use

pooled estimate of the τ^2 (τ^2_{pooled}) [134], i.e. a multivariate version of the pooled estimate for the heterogeneity variance (more technical details are provided in the supplementary material).

We used one of the scores as a common comparator, which would not generally be necessary, but may be easily possible in this MSC setting.

Performance of the considered prognostic scores can be computed from the individual patient data (IPD) directly; this is how we approach the problems having at our disposal a large-scale IPD database. The group results (Stage I) arise from averaging the cohort results, that, in turn, are calculated using the IPD of each cohort. If no IPD are available, instead, there are two possibilities: use published results, or send cohorts code to extract the aggregated performance measures individually. Use of published results requires that comparisons have been reported for more than one score, which in practice may almost never be the case. Sending code to obtain aggregated measures may be an optimal approach in cases where no large-scale collaboration exists, and published results are not detailed enough.

We used all-cause mortality as outcome. Apart from being clinically relevant, mortality is the easiest outcome that we can expect to evaluate in a cohort, with a hard definition. This makes it easier to reduce the problems related to miss-classification or missingness of the outcome [160,161].

Given the patterns of missing data (in general, the variables are completely or almost completely missing or not missing at all) a sensitivity analysis performed after multiple imputation is providing similar results to the analysis without imputation (a comparison is provided in the supplementary material). Analogously, a sensitivity analysis using heterogeneity group by group gives similar results than using a pooled heterogeneity (here recommended because of the network structure; more details are available in the supplementary material).

There are a few limitations to this approach to MSC. Although the analysis can be implemented as outlined by Lu et al [134] (Appendix A), creating an input dataset in a spreadsheet may be less than straightforward. We have therefore provided example R code to convert a dataset of prognostic scores to a MSC meta-analysis, without first making a table of cohort-level summary statistics, as is often performed. We note however that such a dataset including a column for each cell of the variance-covariance matrices could be analysed using mvmeta in Stata. Creating that kind of summary dataset might be useful to go along with the network meta set of commands in Stata [162]. We focused on implementing this approach starting from the raw prognostic scores from individual patients, which had been calculated using raw data from the international collaboration [40].

Conclusions

In summary, we have adapted methodology from network meta-analysis to compare prognostic scores from individual patient data across different cohorts. This approach permits concurrent external validation of the scores in a consistent analysis explicitly including correlations between the scores on a cohort level. Estimates of differences in performance can be estimated for the entire network, as well as for both direct and indirect comparisons of scores. Results of the MSC analysis can be interpreted in a manner similar to that of the usual network meta-analysis, regardless of the performance measure used. Our application to prognostic scores showed that the ADO and updated BODE scores have the best discriminative performance to predict mortality for patients with COPD. The meta-analysis could also be

repeated for a number of different performance measures in order to describe multiple facets of the prognostic scores (e.g. discrimination and calibration [1]) or using reclassification methods (like the net reclassification index, NRI [163]) or to aid in the interpretation of the results. Development of clearer data input formats as well as more automated would provide opportunities for further methodological research in MSC meta-analysis.

Abbreviations: MSC: multiple score comparison; COPD: chronic obstructive pulmonary disease; IPD: individual patient data.

DECLARATIONS

Supplementary material
Supplementary material is available.

Ethics approval and consent to participate
Each individual cohort study was approved by the ethics committee of reference. Details are included in the paper related to the 3CIA collaboration and in the individual studies.

Consent for publication
Not applicable.

Availability of data and material
The dataset used during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Funding
Financial support specific to the current paper came from the University of Zurich (Zurich, Switzerland). The funding of each cohort has been described in detail in the references.

Authors' contributions
SRH drafted the manuscript. BG and SRH performed the statistical analysis. All the authors conceived and designed the study, and performed a critical revision of the manuscript. JBS and MAP supervised the study.

Acknowledgements
We would like to thank Hélène Aschmann, Henock Yebyo and Rafael Sauter (all University of Zurich, Zurich, Switzerland) for their comments.

References

1. Steyerberg EW. *Clinical Prediction Models*. (Gail M, Krickeberg K, Sarnet J, Tsiatis A, Wong W, eds.). Springer - Statistics for Biology and Health - ISBN: 978-1-4419-2648-7; 2010.

2. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: A meta-analysis. *Psychol Assess*. 2000;12(1):19-30. doi:10.1037//1040-3590.12.1.19.

3. Collins GS. The BMJ Gary Collins : Opening up multivariable prediction models. Bmj.

4. Van Gorp MJ, Steyerberg EW, Van Der Graaf Y. Decision Guidelines for Prophylactic Replacement of Bj??rk-Shiley Convexo-Concave Heart Valves: Impact on Clinical Practice. *Circulation*. 2004;109(17):2092-2096. doi:10.1161/01.CIR.0000125853.51637.C8.

5. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation*. 2014;129(25 SUPPL. 1). doi:10.1161/01.cir.0000437738.63853.7a.

6. World Health Organization. *Prevention of Cardiovascular Disease.*; 2007. doi:10.1001/jama.2011.1668.

7. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42(2):517-584. doi:10.1161/STR.0b013e3181fcb238.

8. Expert Panel on Detection, Evaluation and T of HBC in A. Executive summary of the third report of the national cholesterol education program Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Jama*. 2001;285(19):2486-2497. doi:10.1001/jama.285.19.2486.

9. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38. doi:10.1186/1472-6947-6-38.

10. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J Cancer*. 1982;45(3):361-366. doi:10.1038/bjc.1982.62.

11. Di M-Y, Liu H, Yang Z-Y, Bonis PAL, Tang J-L, Lau J. Prediction Models of Mortality in Acute Pancreatitis in Adults. *Ann Intern Med*. 2016;(38). doi:10.7326/M16-0650.

12. Herman WH. Predicting risk for diabetes: Choosing (or building) the right model. *Ann Intern Med*. 2009;150(11):812-814. doi:10.7326/0003-4819-150-11-200906020-00010.

13. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission A Systematic Review. *JAMA*. 2015;306(15).

14. Ma D, Watson L, Ja R, et al. Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes (Review) Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes. 2011;(11). doi:10.1002/14651858.CD004902.pub4.Copyright.

15. Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. *BMJ*. 2016;354(19):i4338. doi:10.1136/bmj.i4338.

16. Wilson PWF, D'Agostino RB, Levy D, Belanger a. M, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation*. 1998;97(18):1837-1847. doi:10.1161/01.CIR.97.18.1837.

17. Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. *Thorax*. 2014;69(7):616-622. doi:10.1136/thoraxjnl-2013-204155.

18. Justice AC, Covinsky KE, Berlin J a. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-524. doi:199903160-00009 [pii].

19. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353(i2416):11. doi:10.1136/bmj.i2416.

20. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? *Bmj*. 1995;311(7019):1539-1541. doi:10.1136/bmj.311.7019.1539.

21. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007.

22. Collins GS, de Groot J a, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40. doi:10.1186/1471-2288-14-40.

23. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research 1: what , why , and how ? *BMJ*. 2009;338(b375):1317-1320. doi:10.1136/bmj.b375.

24. Steyerberg EW, Moons KGM, Windt DA Van Der, et al. Prognosis Research Strategy (PROGRESS) 3 : Prognostic Model Research. *PLoS Med*. 2013;10(2 (e1001381)):9. doi:10.1371/jour-.

25. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247.

26. Bleeker S., Moll H., Steyerberg EW, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol*. 2003;56(9):826-832. doi:10.1016/S0895-4356(03)00207-5.

27. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research 3: validating a prognostic model. *BMJ*. 2009;338(b605):1432-1435. doi:10.1136/bmj.b605.

28. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.

29. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586. doi:10.1002/sim.1844.

30. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method to update prediction models. *Stat Med*. 2014;Under revi(October).

doi:10.1002/sim.7179.

31. Collins GS, Moons KGM. Comparing risk prediction models Should be routine when deriving a new model for the same purpose. *BMJ*. 2012;3186(May):1-2. doi:10.1136/bmj.e3186.

32. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):93-101. doi:10.1002/bjs.9723.

33. Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM. Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLoS Med*. 2015;12(10):1-12. doi:10.1371/journal.pmed.1001886.

34. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Assessing risk prediction models using individual participant data from multiple studies. *Am J Epidemiol*. 2014;179(5):621-632. doi:10.1093/aje/kwt298.

35. Debray TP a, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med 2013;32*. 2013;32:3158-3180. doi:10.1002/sim.5732.

36. Broeze KA, Opmeer BC, Bachmann LM, et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol*. 2009;9:22. doi:10.1186/1471-2288-9-22.

37. Ahmed I, Debray TP a, Moons KGM, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. *BMC Med Res Methodol*. 2014;14(1):3. doi:10.1186/1471-2288-14-3.

38. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. *Br Med J*. 2015;350(1088):1-9. doi:10.1136/bmj.h1088.

39. Siebeling L, ter Riet G, van der Wal WM, et al. ICE COLD ERIC--International collaborative effort on chronic obstructive lung disease: exacerbation risk index cohorts--study protocol for an international COPD cohort study. *BMC Pulm Med*. 2009;9:15. doi:10.1186/1471-2466-9-15.

40. Soriano JB, Lamprecht B, Ramírez AS, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. *Lancet Respir Med*. 2015;3(6):443-450. doi:10.1016/S2213-2600(15)00157-5.

41. Siebeling L, Puhan MA, Muggensturm P, Zoller M, Ter Riet G. Characteristics of Dutch and Swiss primary care COPD patients - baseline data of the ICE COLD ERIC study. *Clin Epidemiol*. 2011;3:273-283. doi:10.2147/CLEP.S24818.

42. Krumholz HM. Why data sharing should be the expected norm. *Bmj*. 2015;350(feb05 6):h599-h599. doi:10.1136/bmj.h599.

43. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *Bmj*. 2016;353(i3140):11. doi:10.1136/bmj.i3140.

44. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data

in meta-analyses using individual participant data: a database survey. *Bmj*. 2012;344(January):1-10. doi:10.1136/bmj.d7762.

45. Moons KGM, Schuit E. Prediction of cardiovascular disease worldwide. *Lancet Diabetes Endocrinol*. 2015;3(5):309-310. doi:10.1016/S2213-8587(15)00002-9.

46. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi:10.1016/j.jclinepi.2014.06.018.

47. Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: The COCOMICS study. *Eur Respir J*. 2013;42(2):323-332. doi:10.1183/09031936.00121012.

48. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol*. 1999;28(1):1-9. doi:10.1093/ije/28.1.1.

49. Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. *Heal (San Fr*. 2001;24(2):218-230.

50. Decramer M, Vogelmeier C, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obrstructive Pulmonary Disease (updated 2017). *Glob Initiat Chronic Obstr Lung Dis*. 2011.

51. Celli BR, MacNee W, Agusti A, et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J*. 2004;23(6):932-946. doi:10.1183/09031936.04.00014304.

52. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP.

53. Miravitlles M, Soler-Cataluña JJ, Calle M, et al. Spanish COPD Guidelines (GesEPOC): pharmacological treatment of stable COPD. Spanish Society of Pulmonology and Thoracic Surgery. *Arch Bronconeumol*. 2012;48(7):247-257. doi:10.1016/j.arbres.2012.04.001.

54. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(10):1005-1012. doi:10.1056/NEJMoa021322.

55. Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open*. 2012;2(6):1-10. doi:10.1136/bmjopen-2012-002152.

56. Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet*. 2009;374(9691):704-711. doi:10.1016/S0140-6736(09)61301-5.

57. Soler-Cataluña JJ, Martinez-Garcia MA, Sanchez LS, Tordera MP, Sanchez PR. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. *Respir Med*. 2009;103(5):692-699. doi:http://dx.doi.org/10.1016/j.rmed.2008.12.005.

58. Esteban C, Quintana JM, Aburto M, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. *Respir Med*. 2011;105(11):1662-1670. doi:http://dx.doi.org/10.1016/j.rmed.2011.05.005.

59. Kostianev SS, Hodgev VA, Iluchev DH. Multidimensional system for assessment of COPD patients. Comparison with BODE index. *Folia Medica (Plovdiv)*. 2008;50(4):29-38.

60. Briggs A, Spencer M, Wang H, Mannino D, Sin DD. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. *Arch Intern Med*. 2008;168(1):71-79. doi:http://dx.doi.org/10.1001/archinternmed.2007.37.

61. Schembri S, Anderson W, Morant S, et al. A predictive model of hospitalisation and death from chronic obstructive pulmonary disease. *Respir Med*. 2009;103(10):1461-1467. doi:http://dx.doi.org/10.1016/j.rmed.2009.04.021.

62. Siebeling L, Musoro JZ, Geskus RB, et al. Prediction of COPD-specific health-related quality of life in primary care COPD patients: a prospective cohort study. *npj Prim Care Respir Med*. 2014;24(14060):7. doi:10.1038/npjpcrm.2014.60.

63. Esteban C, Quintana JM, Aburto M, Moraza J, Capelastegui A. A simple score for assessing stable chronic obstructive pulmonary disease. *QJM*. 2006;99(11):751-759. doi:10.1093/qjmed/hcl110.

64. Boeck L, Soriano JB, Brusse-Keizer M, et al. Prognostic assessment in COPD without lung function: The B-AE-D indices. *Eur Respir J*. 2016;47(6):1635-1644. doi:10.1183/13993003.01485-2015.

65. Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. *Chest*. 2005;127(6):1890-1897. doi:10.1378/chest.127.6.1890.

66. Cardoso F, Tufanin AT, Colucci M, Nascimento O, Jardim JR. Replacement of the 6-min walk test with maximal oxygen consumption in the BODE index applied to patients with COPD: An equivalency study. *Chest*. 2007;132(2):477-482. doi:10.1378/chest.07-0435.

67. Williams JEA, Green RH, Warrington V, Steiner MC, Morgan MDL, Singh SJ. Development of the i-BODE: Validation of the incremental shuttle walking test within the BODE index. *Respir Med*. 2012;106(3):390-396. doi:10.1016/j.rmed.2011.09.005.

68. Almagro P, Barreiro B, Ochoa de Echaguen A, et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. *Respiration*. 2006;73(3):311-317. doi:10.1159/000088092.

69. Thomsen M, Ingebrigtsen TS, Marott JL, et al. Inflammatory biomarkers and exacerbations in chronic obstructive pulmonary disease. *JAMA*. 2013;309(22):2353-2361. doi:http://dx.doi.org/10.1001/jama.2013.5732.

70. Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. *Am J Respir Crit Care Med*. 2009;180(12):1189-1195. doi:http://dx.doi.org/10.1164/rccm.200902-0271OC.

71. Müllerova H, Maselli DJ, Locantore N, et al. Hospitalized Exacerbations of COPD. *CHEST J*. 2015;147(4):999-1007. doi:10.1378/chest.14-0655.

72. Almagro P, Soriano JB, Cabrera FJ, et al. Comorbidities and Short-term Prognosis in Patients

Hospitalized for Acute Exacerbation of COPD. *Chest*. 2012;142(5):1126-1133.

73. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2010;363(11):1128-1138. doi:10.1056/NEJMoa0909883.

74. Moberg M, Vestbo J, Martinez G, et al. Validation of the i-BODE index as a predictor of hospitalization and mortality in patients with COPD Participating in pulmonary rehabilitation. *Copd J Chronic Obstr Pulm Dis*. 2014;11(4):381-387. doi:http://dx.doi.org/10.3109/15412555.2013.836171.

75. Marin JM, Carrizo SJ, Casanova C, et al. Prediction of risk of COPD exacerbations by the BODE index. *Respir Med*. 2009;103(3):373-378. doi:http://dx.doi.org/10.1016/j.rmed.2008.10.004.

76. Ko FW, Tam W, Tung AH, et al. A longitudinal study of serial BODE indices in predicting mortality and readmissions for COPD. *Respir Med*. 2011;105(2):266-273. doi:http://dx.doi.org/10.1016/j.rmed.2010.06.022.

77. Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. *Am J Med Sci*. 2010;339(1):10-14. doi:http://dx.doi.org/10.1097/MAJ.0b013e3181bb8111.

78. Brusse-Keizer M, van der Palen J, van der Valk P, Hendrix R, Kerstjens H. Clinical predictors of exacerbation frequency in chronic obstructive pulmonary disease. *Clin Respir J*. 2011;5(4):227-234. doi:10.1111/j.1752-699X.2010.00234.x.

79. Takahashi T, Muro S, Tanabe N, et al. Relationship between periodontitis-related antibody and frequent exacerbations in chronic obstructive pulmonary disease. *PLoS One*. 2012;7(7):e40570. doi:10.1371/journal.pone.0040570.

80. Azarisman MS, Fauzi MA, Faizal MP, Azami Z, Roslina AM, Roslan H. The SAFE (SGRQ score, air-flow limitation and exercise tolerance) Index: a new composite score for the stratification of severity in chronic obstructive pulmonary disease. *Postgrad Med J*. 2007;83(981):492-497.

81. Motegi T, Jones RC, Ishii T, et al. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int J Copd*. 2013;8:259-271. doi:http://dx.doi.org/10.2147/COPD.S42769.

82. Gudmundsson G, Gislason T, Janson C, et al. Risk factors for rehospitalisation in COPD: role of health status, anxiety and depression. *Eur Respir J*. 2005;26(3):414-419. doi:10.1183/09031936.05.00078504.

83. Niewoehner DE, Lokhnygina Y, Rice K, et al. Risk indexes for exacerbations and hospitalizations due to COPD. *Chest*. 2007;131(1):20-28. doi:10.1378/chest.06-1316.

84. Jacob A, Laurin C, Boudreau M, Lavoie KL, Bacon SL. The impact of body mass index on chronic obstructive pulmonary disease exacerbations. *Can Respir J*. 2013;18(2):237-242.

85. Suetomo M, Kawayama T, Kinoshita T, et al. COPD assessment tests scores are associated with exacerbated chronic obstructive pulmonary disease in Japanese patients. *Respir Investig*. 2014;52:288-295.

86. Parshall MB, Mapel DW, Rice L, Williams A, O'Reilly J. Predictive validity of short-form health

survey 36 items scales for chronic obstructive pulmonary disease exacerbation. *Hear Lung J Acute Crit Care*. 2008;37(5):356-365.

87. Bertens LC, Reitsma JB, Moons KGM, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Copd*. 2013;8:493-499. doi:http://dx.doi.org/10.2147/COPD.S49609.

88. Garcia-Aymerich J, Farrero E, Félez M a, Izquierdo J, Marrades RM, Antó JM. Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. *Thorax*. 2003;58(2):100-105. doi:10.1136/thorax.58.2.100.

89. Chen Y. Readmission in Taiwan. *West J Nurs Res*. 2006;28(1):105-124.

90. Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2014;11(2):149-157. doi:http://dx.doi.org/10.1513/AnnalsATS.201307-243OC.

91. Omachi TA, Yelin EH, Katz PP, Blanc PD, Eisner MD. The COPD severity score: a dynamic prediction tool for health-care utilization. *Copd J Chronic Obstr Pulm Dis*. 2008;5(6):339-346. doi:http://dx.doi.org/10.1080/15412550802522700.

92. Lee SD, Huang MS, Kang J, et al. The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. *Respir Med*. 2014;108(4):600-608.

93. Amalakuhan B, Kiljanek L, Parvathaneni A, Hester M, Cheriyath P, Fischman D. A prediction model for COPD readmissions: catching up, catching our breath, and improving a national problem. *J Community Hosp Intern Med Perspect*. 2012;2(1):9915. doi:http://dx.doi.org/10.3402/jchimp.v2i1.9915.

94. Ong KC, Earnest A, Lu SJ. A multidimensional grading system (BODE index) as predictor of hospitalization for COPD. *Chest*. 2005;128(6):3810-3816.

95. Echave-Sustaeta J, Comeche Casanova L, Garcia Lujan R, Sayas Catalan J, Gomez de la Camara A, Lopez Encuentra A. Prognosis following acute exacerbation of COPD treated with non-invasive mechanical ventilation. *Arch Bronconeumol*. 2010;46(8):405-410. doi:http://dx.doi.org/10.1016/j.arbres.2010.03.004.

96. Decramer M. Management of COPD according to guidelines. A national survey among Belgian physicians. *Monaldi Arch Chest Dis*. 2003.

97. Rutschmann OT. Knowledge of guidelines for the management of COPD: a survey of primary care physicians. *Respir Med*. 2004.

98. Puhan MA, Zoller M, Ter Riet G. COPD : more than respiratory (comment). *Lancet (comment)*. 2008;371:26-27.

99. Fritsch K, Jacot M-L, Klarer A, et al. Adherence to the Swiss guidelines for management of COPD: experience of a Swiss teaching hospital. *Swiss Med Wkly*. 2005;135(7-8):116-121. doi:2005/07/smw-10844.

100. Calverley PMA, Anderson JA, Celli B, et al. Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2007;356(8):775-789.

101. Vestbo J, Anderson JA, Brook RD, et al. Fluticasone furoate and vilanterol and survival in chronic obstructive pulmonary disease with heightened cardiovascular risk (SUMMIT): A double-blind randomised controlled trial. *Lancet*. 2016;387(10030):1817-1826. doi:10.1016/S0140-6736(16)30069-1.

102. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet*. 2007;370(9589):797-799. doi:10.1016/S0140-6736(07)61383-X.

103. Fabbri LM, Luppi F, Beghé B, Rabe KF. Complex chronic comorbidities of COPD. *Eur Respir J*. 2008;31(1):204-212. doi:10.1183/09031936.00114307.

104. Hamburg MA, Collins FS. The Path to Personalized Medicine. *N Engl J Med*. 2010;363(4):301-304. doi:10.1056/NEJMp1002530.

105. Weston AD, Hood L. Systems Biology , Proteomics , and the Future of Health Care : Toward Predictive , Preventative , and Personalized Medicine Introduction : Paradigm Changes in Health Care. *J Proteome Res*. 2004;3(2):179-196. doi:10.1021/pr0499693.

106. Agusti A. The path to personalised medicine in COPD. *Thorax*. 2014;0(4):1-8. doi:10.1056/NEJMp1002530.

107. Agusti A, Sobradillo P, Celli BR. Addressing the complexity of chronic obstructive pulmonary disease: from phenotypes and biomarkers to scale-free networks, systems biology, and P4 medicine. *Am J Respir Crit Care Med*. 2011;183(9):1129-1137. doi:10.1164/rccm.201009-1414PP.

108. Decramer M, Vestbo J. *Global Initiative for Chronic Obstructive Lung Disease (Updated 2014)*.; 2014.

109. Puhan MA, Bachmann LM, Kleijnen J, Ter Riet G, Kessels AG. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Med*. 2009;7:2. doi:10.1186/1741-7015-7-2.

110. Hingorani ADA, Windt DVD a Van Der, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj*. 2013;346(February 2013):e5793. doi:10.1136/bmj.e5793.

111. Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. *BMC Med*. 2013;11(1):7. doi:10.1186/1741-7015-11-7.

112. *Systematic Reviews (CRD Guidance)*. Center for Reviews and Dissemination, University of York; 2009.

113. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *Annu Intern Med*. 2009;151(4):264-269. doi:10.1371/journal.pmed1000097.

114. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100. doi:10.1371/journal.pmed.1000100.

115. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann*

Intern Med. 2015;162(1):55-64. doi:10.7326/M14-0697.

116. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

117. Cochrane Statistical Methods group and Cochrane Bias Methods Group. *Assessing Risk of Bias in Included Studies*.; 2011.

118. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-781.

119. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690. doi:10.1136/heartjnl-2011-301246.

120. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research 4: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(7709):1487-1490. doi:10.1136/bmj.b606.

121. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61(11):1085-1094. doi:10.1016/j.jclinepi.2008.04.008.

122. Alonzo TA. Clinical Prediction Models: A Practical Approach to Development, Validation , and Updating (Book Review). *AJE (American J Epidemiol*. 2009;170(4):90033. doi:10.1093/aje/kwp129.

123. Khalili D, Hadaegh F, Soori H, Steyerberg EW, Bozorgmanesh M, Azizi F. Clinical Usefulness of the Framingham Cardiovascular Risk Profile Beyond Its Statistical Performance The Tehran Lipid and Glucose Study. *AJE (American J Epidemiol*. 2012;176(3):177-186. doi:10.1093/aje/kws204.

124. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Risk Factor Models for Neurodevelopmental Outcomes in Children Born Very Preterm or With Very Low Birth Weight : A Systematic Review of Methodology and Reporting. *AJE (American J Epidemiol*. 2017;185(7):601-612. doi:10.1093/aje/kww135.

125. Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF. Predicting Risk of Type 2 Diabetes Mellitus with Genetic Risk Models on the Basis of Established Genome-wide Association Markers : A Systematic Review. *AJE (American J Epidemiol*. 2013;178(8):1197-1207. doi:10.1093/aje/kwt123.

126. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev*. 2017;26(143):1-13. doi:10.1183/16000617.0061-2016.

127. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *AJE (American J Epidemiol*. 2006;163(9):783-789. doi:10.1093/aje/kwj093.

128. Collins GS, Moons KGM. Comparing risk prediction models. *Br Med J*. 2012;344(may24 2):e3186-e3186. doi:10.1136/bmj.e3186.

129. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*.

1996;15(24):2733-2749. doi:10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.

130. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(November 2001):2313-2324. doi:10.1002/sim.1201.

131. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105-3124. doi:10.1002/sim.1875.

132. Lu G, Ades a. E. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *J Am Stat Assoc*. 2006;101(474):447-459. doi:10.1198/016214505000001302.

133. Salanti G, Higgins J, Ades a E, Ioannidis JP a. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008;17(3):279-301. doi:10.1177/0962280207080643.

134. Lu G, Welton NJ, Higgins J, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Res Synth Methods*. 2011;2(1):43-60. doi:10.1002/jrsm.34.

135. Zarin W, Veroniki AA, Nincic V, et al. Erratum to: Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med*. 2017;15(1):61. doi:10.1186/s12916-017-0832-6.

136. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385.

137. Panagiotou OA. Network Meta-Analysis : Evidence Synthesis With Mixed Treatment Comparison (Book Review). *AJE (American J Epidemiol*. 2015;181(4):288-289. doi:10.1093/aje/kwu471.

138. Takwoingi Y. Empirical Evidence of the Importance of Comparative Studies of Diagnostic Test Accuracy. *Ann Intern Med*. 2013;158.

139. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res*. 2015:0962280215592269-. doi:10.1177/0962280215592269.

140. Manolio TA, Weis BK, Cowie CC, et al. New Models for Large Prospective Studies : Is There a Better Way ? *AJE (American J Epidemiol*. 2012;175(June 2010):859-866. doi:10.1093/aje/kwr453.

141. Franchini a. J, Dias S, Ades a. E, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. *Res Synth Methods*. 2012;3(2):142-160. doi:10.1002/jrsm.1049.

142. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(March):80-97. doi:10.1002/jrsm.1037.

143. Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology*. 1983;148:839-843.

144. Borenstein M, Hedges LV., Higgins J, R. RH. *Introduction to Meta-Analysis*. Wiley; 2011.

145. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188.

doi:10.1016/0197-2456(86)90046-2.

146. Nikolakopoulou A, Mavridis D, Salanti G. How to interpret meta-analysis models: fixed effect and random effects meta-analyses. *Evid Based Ment Health*. 2014;17(2).

147. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health*. 2014;17(2):53-57.

148. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*. 2012;12(1):152. doi:10.1186/1471-2288-12-152.

149. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods*. 2013;4(4):291-323. doi:10.1002/jrsm.1085.

150. Cooper NJ, Sutton AJ, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28(15):1982-1998. doi:10.1002/sim.

151. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med*. 2013;11(1):159. doi:10.1186/1741-7015-11-159.

152. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and Technical Challenges in Network Meta-analysis. *Ann Intern Med*. 2013;159:130-137. doi:10.7326/0003-4819-159-2-201307160-00008.

153. Efthimiou O, Debray TP a, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;(November 2014). doi:10.1002/jrsm.1195.

154. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins J. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014;9(7):e99682. doi:10.1371/journal.pone.0099682.

155. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971-980. doi:10.1093/aje/kwq223.

156. Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-1573. doi:10.1002/sim.1187.

157. Dias S, Welton NJ, Caldwell DM, Ades a. E. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(November 2008):932-944. doi:10.1002/sim.3767.

158. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley Serie in Probability and Mathematical Statistics; 1987. doi:10.1002/9780470316696.

159. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol*. 2010;10(1):112. doi:10.1186/1471-2288-10-112.

160. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for Misclassified Outcomes in Binary Regression Models Using Multiple Imputation With Internal Validation Data. *AJE (American J Epidemiol*. 2013;177(9):904-912. doi:10.1093/aje/kws340.

161. Groenwold RHH, Donders ART, Roes KCB, Harrell FE, Moons KGM. Dealing With Missing Outcome Data in Randomized Trials and Observational Studies (Practice of Epidemiology). *AJE (American J Epidemiol*. 2012;175(3):210-217. doi:10.1093/aje/kwr302.

162. White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985. doi:The Stata Journal.

163. Pepe MS. Problems With Risk Reclassification Methods for Evaluating Prediction Models. *AJE (American J Epidemiol*. 2011;173(11):1327-1335. doi:10.1093/aje/kwr013.

164. Harrell FE. *Regression Modelling Strategies*. (Bickel P, Diggle P, Feinberg SE, Gather U, Olkin I, Zeger S, eds.). Springer - Statistics for Biology and Health; ISBN: 978-3-319-19424-0; 2015.

165. Laupacis A, Wells G, Richardson W. S, Tugwell P. Users' guides to the medical literature: V. How to use an article about prognosis. *Jama*. 1994;272(3):234-237. doi:doi:10.1001/jama.1994.03520030076032.

166. Puhan MA, Yu T, Stegeman I, Varadhan R, Singh S, Boyd CM. Benefit-harm analysis and charts for individualized and preference-sensitive prevention: example of low dose aspirin for primary prevention of cardiovascular disease and cancer. *BMC Med*. 2015;13(250):11. doi:10.1186/s12916-015-0493-2.

167. Hayes DF, Markus HS, Leslie RD, Topol EJ. Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC Med*. 2014;12(37):8. doi:10.1186/1741-7015-12-37.

168. Nickel CH, Bingisser R, Morgenthaler NG. The role of copeptin as a diagnostic and prognostic biomarker for risk stratification in the emergency department. *BMC Med*. 2012;10(7). doi:10.1186/1741-7015-10-7.

169. Vuong K, McGeechan K, Armstrong BK, Cust AE. Risk Prediction Models for Incident Primary Cutaneous Melanoma: A Systematic Review. *JAMA Dermatology*. 2014;150(4):434. doi:10.1001/jamadermatol.2013.8890.

170. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA - J Am Med Assoc*. 2011;306(15):1688-1698.

171. Yourman LC, Lee SJ, Schonberg M a, Widera EW, Smith AK. Prognostic Indices for Older Adults: A Systematic Review. *JAMA*. 2012;307(2):182-192.

172. Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9(103):14. doi:10.1186/1741-7015-9-103.

173. Esteban C, Quintana JM, Moraza J, et al. BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice? *BMC Med*. 2010;8:28. doi:10.1186/1741-7015-8-28.

174. Quintana JM, Esteban C, Unzurrunzaga A, et al. Predictive score for mortality in patients with COPD exacerbations attending hospital emergency departments. *BMC Med*. 2014;12:66.

doi:http://dx.doi.org/10.1186/1741-7015-12-66.

175. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research 2: Developing a prognostic model. *BMJ (Research Methods Reporting)*. 2009;338(june):1373-1377. doi:10.1136/bmj.b604.

176. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med*. 2010;8(21):11. doi:1741-7015-8-21 [pii] 10.1186/1741-7015-8-21.

177. Hemingway H. Ten steps towards improving prognosis research Problems with prognosis research. *BMJ*. 2014;4184(December 2009):1-10.

178. Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med*. 2016;14(104):15. doi:10.1186/s12916-016-0650-2.

179. Thangaratinam S, Allotey J, Marlin N, et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med*. 2017;15(68):11. doi:10.1186/s12916-017-0827-3.

180. Haile SR, Guerra B, Soriano JB, Puhan MA. Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores. *Stat Methods Med Res*. 2017;17(172):1-12. doi:(submitted).

181. Kessels AG, Riet G, Puhan MA, Kleijnen J, Bachmann LM, Minder C. A simple regression model for network meta-analysis. *OA Epidemiol*. 2013;1(1):1-8.

182. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med*. 2011;9(79):5. doi:10.1186/1741-7015-9-79.

183. Salanti G, Marinho V, Higgins J. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol*. 2009;62(8):857-864. doi:10.1016/j.jclinepi.2008.10.001.

184. Sauter R, Held L. Network meta-analysis with integrated nested Laplace approximations. *Biometrical J*. 2015;57(6):1038-1050. doi:10.1002/bimj.201400163.

185. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating (ROC) Curvel Characteristic. *Radiology*. 1982;143:29-36.

186. Caldwell DM, Ades a E, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj*. 2005;331(7521):897-900. doi:10.1136/bmj.331.7521.897.

187. Mills EJ, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. *Jama*. 2012;308(12):1246-1253. doi:10.1001/2012.jama.11228.

188. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *Bmj*. 2017;356(i6460):11. doi:10.1136/bmj.i6460.

189. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting

results from multiple-treatment meta-analysis: An overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-171. doi:10.1016/j.jclinepi.2010.03.016.

190. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15(58):9. doi:10.1186/s12874-015-0060-8.

191. Ransohoff DF, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-930.

192. Ioannidis JP a, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. *JAMA - J Am Med Assoc*. 2010;303(16):1646-1647. doi:10.1001/jama.2010.503.

193. Divo M, Cote C, De Torres JP, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155-161. doi:10.1164/rccm.201201-0034OC.

194. Reilly BM, Evans AT, Schaider JJ, et al. Impact of a Clinical Decision Rule in the Emergency Department. *JAMA - J Am Med Assoc*. 2002;288(3):342-350.

195. McGinn TG, Guyatt GH, Wyer PC, David Naylor C, Stiell IG, Scott Richardson W. Users' Guides to the Medical Literature. *J Am Med Assoc*. 2000;284(1):79-84. doi:10.1001/jama.284.1.79.

196. Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA - J Am Med Assoc*. 2009;302(21):2345-2352. doi:10.1001/jama.2009.1757.

197. Wedzicha JA, Brill SE, Allinson JP, Donaldson GC. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. *BMC Med*. 2013;11(181):10. doi:10.1186/1741-7015-11-181.

198. Parkes G. Using risk scores with patients. *Bmj*. 2010;340(jun22 3):c3327-c3327. doi:10.1136/bmj.c3327.

199. Steyerberg EW, Bleeker S., Moll H a, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441-447. doi:10.1016/S0895-4356(03)00047-7.

200. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005.

201. Reilly BM, Evans AT. Academia and clinic. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.

202. Held U, Garcia-Aymerich J, Ter Riet G, Moons KGM, Puhan MA. Missing variables in risk prediction models. *AJE (American J Epidemiol*. 2012. doi:10.1002/elan.

203. Harrell FE. *Regression Modeling Strategies*. Springer; 2015.

204. Snell KIE, Hua H, Debray TPA, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2016;69:40-50. doi:10.1016/j.jclinepi.2015.05.009.

205. Blattenberger G, Lad F. Separating the Brier Score into Calibration and Refinement Components :

A Graphical Exposition Author (s): Gail Blattenberger and Frank Lad Source : The American Statistician , Vol . 39 , No . 1 (Feb ., 1985), pp . 26-32 Published by : Taylor & Franci. *Am Stat*. 1985;39(1):26-32.

206. Murphy AH. A New Vector Partition of the Probability Score. *J Appl Metereology*. 1973;12:595-600.

CHAPTER 4: Application of MSC Meta-Analysis

Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores

Beniamino Guerra, Dr. Sarah R. Haile, Prof. Bernd Lamprecht, Dr. Ana S. Ramírez, Dr. Pablo Martinez-Camblor, Dr. Bernhard Kaiser, Dr. Inmaculada Alfageme, Dr. Pere Almagro, Dr. Ciro Casanova, Dr. Cristóbal Esteban-González, Dr. Juan J. Soler-Cataluña, Dr. Juan P. de Torres, Prof. Marc Miravitlles, Prof. Bartolome R. Celli, Dr. Jose M. Marin, Dr. Gerben ter Riet, Dr. Patricia Sobradillo-Ecenarro, Prof. Peter Lange, Prof. Judith Garcia-Aymerich, Prof. Josep M. Antó-Boqué, Dr. Alice M. Turner, Dr. Meilan K. Han, Dr. Arnulf Langhammer, Dr. Linda Leivseth, Dr. Per Bakke, Dr. Ane Johannessen, Dr. Oga Toru, Dr. Borja Cosío, Dr. Julio Ancochea-Bermúdez, Dr. Andres Echazarreta, Prof. Nicolas Roche, Prof. Pierre-Régis Burgel, Prof. Don D. Sin, Prof. Joan B. Soriano, Prof. Milo A. Puhan, for the 3CIA collaboration

Guerra et al. BMC Medicine (2018) 16:33

Complete List of Authors (with affiliation and email):

1. Guerra, Beniamino; Universitat Zurich, Epidemiology, Biostatistics and Prevention Institute (Beniamino.guerra@uzh.ch)
2. Haile, Sarah; Universitat Zurich, Epidemiology, Biostatistics and Prevention Institute (sarah.haile@uzh.ch)
3. Lamprecht, Bernd; Kepler Universitatsklinikum GmbH, Department of Pulmonary Medicine; Johannes Kepler Universitat Linz, Faculty of Medicine (bernd.lamprecht@kepleruniklinikum.at)
4. Ramírez-GarciaLuna, Ana; Universidad Autonoma de San Luis Potosi, Facultad de Medicina UASLP (asr.3cia@gmail.com)
5. Martinez-Camblor, Pablo; Dartmouth College Geisel School of Medicine (pmcamblor@hotmail.com)
6. Kaiser, Bernhard; Paracelsus Medizinische Privatuniversitat, Department of Pulmonary Medicine (bernhardkaiser@gmx.at)
7. Alfageme, Inmaculada; Hospital Universitario de Valme (inma.alfageme.michavila@gmail.com)
8. Almagro, Pedro; Hospital Universitario Mutua de Terrassa, Internal Medicine (19908pam@comb.cat)
9. Casanova, Ciro; Hospital Universitario NS La Candelaria, Pulmonary Department and Research Unit (casanovaciro@gmail.com)
10. Esteban-González, Cristóbal; Hospital Galdakao-Usansolo (cristobal.est@gmail.com)
11. Soler-Cataluña, Juan S.; Hospital Universitari Arnau de Vilanova, Servicio de Neumología (jjsoler@telefonica.net)
12. de Torres, Juan P.; Clinica Universidad de Navarra, Pulmonary Department (jupa65@hotmail.com)
13. Miravitlles, Marc; Hospital Vall d'Hebron, Pneumology Department; CIBER de Enfermedades Respiratorias (CIBERES). Barcelona, Spain (mmiravitlles@vhebron.net)
14. Celli, Bartolome; Brigham and Women's Hospital, Pulmonary and Critical Care Medicine (BCelli@copdnet.org)
15. Marin, Jose; Hospital Universitario Miguel Servet, IISAragon and CIBERES (jmmarint@unizar.es)
16. ter Riet, Gerben; Academic Medical Center, University of Amsterdam, Department of General Practice (g.terriet@amc.uva.nl)
17. Sobradillo-Ecenarro, Patricia; Universidad del Pais Vasco, Universitat de Barcelona (patricia.sobradilloecenarro@osakidetza.eus)
18. Lange, Peter; Department of Public Health, Section of Social Medicine, University of Copenhagen (peter.lange@sund.ku.dk)
19. Garcia-Aymerich, Judith; ISGlobal, Barcelona, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain; (judith.garcia@isglobal.org)
20. Antó-Boqué, Josep M.; Instituto de Salud Global Barcelona, Centre for Research in Environmental Epidemiology (CREAL); Institut Hospital del Mar d'Investigacions Mediques (josepm.anto@isglobal.org)
21. Turner, Alice; University of Birmingham, Institute of Inflammation and Aging (A.M.Turner@bham.ac.uk)

22. Han, Meilan K.; University of Michigan, Division of Pulmonary and Critical Care (mrking@med.umich.edu)
23. Langhammer, Arnulf; Norvegian University of Science and Technology, Department of Public Health and Nursing (arnulf.langhammer@ntnu.no)
24. Leivseth, Linda; Centre for Clinical Documentation and Evaluation, Northern Norway Regional Health Authority (Linda.Leivseth@helse-nord.no)
25. Bakke, Per; University of Bergen, Haukeland University Hospital (Per.Bakke@k2.uib.no)
26. Johannessen, Ane; University of Bergen, Dept. of Global Public Health and Primary Care (ane.johannessen@uib.no)
27. Oga, Toru; Kyoto University, Department of Respiratory Care and Sleep Control Medicine, Graduate School of Medicine (ogato@kuhp.kyoto-u.ac.jp)
28. Cosio, Borja G; Hospital Son Espases-IdISBa-CIBERES , Department of Respiratory Medicine, Palma de Mallorca, Spain (borja.cosio@ssib.es)
29. Ancochea-Bermúdez, Julio; Instituto de Investigación del Hospital Universitario de la Princesa (IISP), Universidad Autónoma de Madrid, Madrid, Spain , Servicio de Neumología (juli119@gmail.com)
30. Echazarreta, Andres; Universidad Nacional de la Plata, Hospital San Juan de Dios de La Plata (aechaza@gmail.com)
31. Roche, Nicolas; Hopitaux Universitaires Paris Centre, Service de Pneumologie AP-HP (nicolas.roche@cch.aphp.fr)
32. Burgel, Pierre; Hopital Cochin; Universite Paris Descartes (pierre-regis.burgel@aphp.fr)
33. Sin, Don D.; University of British Columbia, James Hogg Research Centre (don.sin@hli.ubc.ca)
34. Soriano, B. Joan; Instituto de Investigación del Hospital Universitario de la Princesa (IISP), Universidad Autónoma de Madrid, Servicio de Neumología, Madrid, Spain; and Scientific and Methodological Consultant of SEPAR www.separ.es , Barcelona, Spain (jbsoriano2@gmail.com)
35. Puhan, Milo; Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland; Epidemiology & Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health (miloalan.puhan@uzh.ch)

CHAPTER 4: Application of MSC Meta-Analysis

Abstract

Background: External validations and comparisons of prognostic models or scores are a pre-requisite for their use in routine clinical care but lacking in most medical fields including chronic obstructive pulmonary disease (COPD). Our aim was to externally validate and concurrently compare prognostic scores for 3-year all-cause mortality in mostly multimorbid patients with COPD.

Methods: We relied on 24 cohort studies of the COPD Cohorts Collaborative International Assessment (3CIA) consortium, corresponding to primary, secondary and tertiary care in Europe, the Americas and Japan. They include globally 15762 patients with COPD (1871 deaths and 42203 person-years of follow-up). We used network meta-analysis adapted to multiple score comparison (MSC), following a frequentist two-stage approach; thus, we were able to compare all scores in a single analytical framework accounting for correlations among scores within cohorts. We assessed transitivity, heterogeneity and inconsistency and provided a performance ranking of the prognostic scores.

Results: Depending on data availability, between 2 and 9 prognostic scores could be calculated for each cohort. The BODE score (body mass index, airflow obstruction, dyspnea and exercise capacity) had a median AUC of 0.679 [1st quartile-3rd quartile = 0.655-0.733] across cohorts. The ADO score (age, dyspnea and airflow obstruction) showed the best performance for predicting mortality (difference AUC_{ADO} - AUC_{BODE} = 0.015 [95% confidence interval (CI) = -0.002 to 0.032; p = 0.08) followed by the updated BODE ($AUC_{BODE\ updated}$ - AUC_{BODE} = 0.008 [95% CI = -0.005 to +0.022; p = 0.23). The assumption of transitivity was not violated. Heterogeneity across direct comparisons was small and we did not identify any local or global inconsistency.

Conclusions: Our analyses showed best discriminatory performance for the ADO and updated BODE scores in patients with COPD. A limitation to be addressed in future studies is the extension of MSC network meta-analysis to measures of calibration. MSC network meta-analysis can be applied to prognostic scores in any the medical field to identify the best scores, possibly paving the way for stratified medicine, public health and research.

Keywords: COPD, prognostic scores, large-scale external validation, performance comparison, network meta-analysis

CHAPTER 4: Application of MSC Meta-Analysis

Background

Prognostic scores, commonly based on coefficients from regression models, provide a probability of a certain adverse outcome for an individual over a specified time horizon. Prognostic scores have become increasingly popular over the last two decades [1,23,24,164,165]. They serve multiple purposes such as informing individuals and health care providers about disease and outcome risks, supporting risk-stratified and personalized prevention or treatment decisions, identifying participants for research or adjusting for confounding [111,166–168].

Numerous prognostic models have been developed in various fields of Medicine [169–172]. Just for predicting the risk of cardiovascular disease in the general population, a recent review identified 363 prognostic models or scores [19]. For patients with Chronic Obstructive Pulmonary Disease (COPD), prognostic scores have been developed mostly to predict the risk of death [54–58,60,61,63–67,70,80,173,174], but scores also exist to predict exacerbations [126], or deteriorating of health-related quality of life [62,70].

Major obstacles for using prognostic scores in practice and research are, however, the frequent lack of external validations, comparisons of their predictive performance and assessments of their applicability in practice [23,27,31,43,120,175,176]. Practitioners and researchers are left with uncertainty about which prognostic score to use and may be reluctant to use them at all [177]. Ideally, prognostic scores would be externally validated in several different populations and their performance summarized [178,179]. However, such external validations and concurrent comparisons are rarely performed [47]. In addition, for even more comprehensive comparison, the performance of prognostic scores may be compared indirectly using common comparator scores similar to network meta-analysis (NMA) [142,180–184] of randomized trials.

Our aim was to use Multiple Score Comparison (MSC) in order to externally validate and concurrently compare prognostic scores for 3-year mortality in patients with Chronic Obstructive Pulmonary Disease (COPD).

Methods

We followed a pre-specified study protocol and described the detailed statistical methods elsewhere.[180]

Study design and participants

This study was based on 26 cohort studies of the COPD Cohorts Collaborative International Assessment (3CIA) consortium. Details have been reported elsewhere and summarized in Table 2 [40]. All cohorts were approved by ethics committees and participants gave written informed consent [40]. We included also the PAC-COPD and Copenhagen cohorts in the final database, even if they were used in the large-scale update of the ADO index [55]. We considered this approach reasonable, since they form only small part of the final database but we verified in a sensitivity analysis if they affected the results.

Prognostic scores

Starting from the literature review of two studies [47,62] and searching among their references, Pubmed related articles and through our research network, we identified 19 prognostic scores, of which we included 10 in our analysis (in the appendix the whole list with corresponding scoring rules). The scores (see Table 1 for details) were the BODE (Body mass index, airflow Obstruction, Dyspnea and

severe Exacerbations) [54], updated BODE [56], ADO (Age, Dyspnea, airflow Obstruction; we included in the analysis only the updated ADO index and not the original ADO index [56] because the updated ADO was generated from large-scale external validation; however we will name it simply ADO) [55], eBODE (severe acute exacerbation of COPD plus BODE) [57], BODEx (Body mass index, airflow Obstruction, Dyspnea, severe acute Exacerbation of COPD) [57], DOSE (Dyspnea, Obstruction, Smoking and Exacerbation frequency) [70], SAFE (Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity) [80], B-AE-D (Body-mass index, Acute Exacerbations, Dyspnea; we used the optimized version and not the original B-AE-D score) [64]. The GOLD classification [50,52] and the 2011-2016 GOLD classification (often referred as new GOLD in the recent COPD literature) [50] were also used in the analysis, even if they were not designed for prognostic purposes. Apart from original ADO and original B-AE-D score (in the appendix we evaluated also their performances) the other 7 identified scores from literature were excluded from the analysis, since our database did not include at least one of their predictors or did not include them simultaneously in at least 1 cohort.

Table 1 Scoring rules of prognostic scores to predict mortality in patients with COPD

Score Predictor	GOLD [50,52]	GOLD (2011-2016) [50]	BODE [54]	BODE upd. [56]	ADO [55]	e-BODE [57]	BODEx [57]	DOSE [70]	SAFE [80]	B-AE-D [64]
BMI			0 (>21) 1 (<=21)	0 (>21) 1 (<=21)		0 (>21) 1 (<=21)	0 (>21) 1 (<=21)			0 (>=21) 6 (18.5-21) 9 (<18.5)
FEV1% pred.	0 (>=80%) 1 (50-79%) 2 (30-49%) 3 (<30%)	0 (if FEV1pp>=50 and <=1 exacerbations per year) 2 (otherwise)	0 (>=65%) 1 (50-64%) 2 (36-49%) 3 (<=35)	0 (>=65%) 1 (36-64%) 2 (<=35)	0 (>=81%) 1 (65-60%) 2 (51-64%) 3 (35-50%) 4 (<=35%)	0 (>=65%) 1 (50-64%) 2 (36-49%) 3 (<=35)	0 (>=65%) 1 (50-64%) 2 (36-49%) 3 (<=35)	0 (>=50%) 1 (31-49%) 2 (<=30)	0 (>=80%) 1 (50-79%) 2 (36-49%) 3 (<=35)	
mMRC		0 (if mMRC >=2 and CAT >=10) 1 otherwise	0 (0-1) 1 (2) 2 (3) 3 (4)	0 (0-1) 1 (2) 2 (3) 3 (4)	0 (0) 1 (1-2) 2 (3) 3 (4)	0 (0-1) 1 (2) 2 (3) 3 (4)	0 (0-1) 1 (2) 2 (3) 3 (4)	0 (0-1) 1 (2) 2 (3) 3 (4)		0 (0-2) 6 (3) 10 (4)
6-MWT (m)			0 (>=350) 1 (250-349) 2 (150-249) 3 (<=149)	0 (>=350) 4 (250-349) 7 (150-249) 9 (<=149)		0 (>=350) 1 (250-349) 2 (150-249) 3 (<=149)			0 (>=400) 1 (300-399) 2 (200-299) 3 (<=199)	
Age (y)					0 (40-49) 2 (50-59)					

					4 (60-69) 5 (70-79) 7 (>=80)					
Prev. exacerbation		(See FEV1pp)				0 (0) 1 (1-2) 2 (>2)	0 (0) 1 (1-2) 2 (>2)	0 (0-1) 1 (2-3) 2 (>3)		0 (0) 3 (1) 7 (>=2)
CAT		(See mMRC)								
Smoking								0 (non-smoker) 1(current smoker)		
Quality of life (SGRQ)									0 (<=30) 1 (31-49) 2 (50-64) 3 (>=65)	
Total Score	0-3	0-3	0-10	0-15	0-14	0-12	0-9	0-8	0-9	0-26

Abbreviations: BMI=body-mass index; FEV1% pred.=forced expiratory volume in 1 second percentage predicted; mMRC=modified Medical Research Council dyspnea scale; 6MWT=6-minute walk test; CAT=COPD Assessment Test; SGRQ=Saint George's Respiratory Questionnaire; previous exacerbations are referred to the previous year. GOLD=Global initiative for chronic Obstructive Lung Disease; BODE=Body mass index, airflow Obstruction, Dyspnea and severe Exacerbations; BODE upd.=BODE updated; ADO=Age, Dyspnea, airflow Obstruction (we use in the our analysis the updated version of the ADO score); e-BODE=severe acute exacerbation of COPD plus BODE; BODEx=Body mass index, airflow Obstruction, Dyspnea, severe acute Exacerbation of COPD; DOSE=Dyspnea, Obstruction, Smoking and Exacerbation frequency; SAFE=Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D.=Body-mass index, Acute Exacerbations, Dyspnea (we use the optimized version of the score, introduced in the same paper). Missing cells correspond to variables that do not constitute the score of the correspondent column.

Outcome and performance measure for external validation and comparison of prognostic scores

We evaluated a number of performance measures commonly used to assess the prognostic properties of prediction models and scores.[180] We deemed the Area Under the Curve (AUC) to be the most appropriate performance measure for our purposes, mainly because its range is independent of the data, it is easy to interpret and an analytic formula for its variance is available [185].

Statistical analysis

We followed a pre-specified study protocol. We first performed direct head-to-head comparisons using random-effects meta-analysis, and then examined the network evidence merging all available direct and indirect evidence [186]. We used a novel methodology, i.e. Multiple Score Comparison (MSC) meta-analysis, adapted from multiple treatment comparison network meta-analysis [134,187]. Methodological details are reported in the section "Detailed Methods" in the supplementary material and in a recent paper [180]. R codes are available (provided in the section "R Code for MSC meta-analysis" in the supplementary material).

Direct comparisons (random-effects pairwise meta-analysis)

We directly compared prognostic scores by pairwise random-effects meta-analysis [145,188]. We used forest plots to visually investigate statistical heterogeneity, as well as the I^2 statistic. Such standard meta-analysis has limitations, since it does not take into account the correlations among multiple scores evaluated on the same set of patients [141] and it does not give a clear indication of which prognostic score performs best. Thus, we adopted network meta-analysis, an approach that allowed us to weight and then pool the results coming from different cohorts.

MSC meta-analysis

Methodological details are reported in details in [180]. In brief, we used an example of implementation of network meta-analysis for treatment effectiveness comparison [134], adapting it to our purposes, namely to concurrently externally validate and compare prognostic scores from individual patient data across different cohorts (reference coming soon and proof reading already available). We have explicitly included correlations [141] between the scores on a cohort level. We use a frequentist two-stage meta-regression model, as proposed in [134]:

1. Ordinary meta-analysis (stage I) to obtain the direct estimates for pooled differences in AUC (using the inverse-variance weighted means of the corresponding cohorts). The meta-analyses were done within each group of cohorts where data for the same prognostic scores were available.
2. In this stage II, we merged the estimates for the differences in AUC from the groups of cohorts, looking for the weighted least squares solution to the regression problem equation. Based on the direct estimates and their variances from the first stage, we estimated the pooled differences in AUC that obeyed fundamental consistency equations. Thus in stage II, the Stage I estimates for the differences in AUC were combined across groups of cohorts to give overall performance estimates for the entire network.

In order to provide a ranking of the scores, we used a frequentist version of the SUCRA (Surface Under the Cumulative Ranking Curve; [189,190]) score showing the likelihood of score to be better than any other score and summarizing relative performances and confidence intervals.

The last steps were to ensure that the heterogeneity, transitivity and consistency assumptions were met.[142] Heterogeneity in the MSC analysis was evaluated by the pooled heterogeneity variance among groups (τ^2_{pooled}). We assessed "transitivity" through ANOVA tests. Thus we assessed the comparability of the cohorts across whom the predictive performance of a score may vary because of a "spectrum effect"[191] or "case mix" [43,46,155]. We also assessed consistency [142] between direct evidence and MSC meta-analysis estimates using Q likelihood-ratio test statistic to evaluate the global consistency and analysis of residuals and leverages to evaluate the local consistency [134]. For more details, see "Detailed Methods" in the supplementary material and [180]

Handling of missing data

If a variable was missing for >30% of the patients we discarded the specific variable for that particular specific cohort, since the effects of such predictors could be generally distrusted [1]. Otherwise

we performed multiple imputation with chained equations (the analysis of the patterns of missingness allowed us to consider the missing data missing completely at random apart from the dependence on the cohort) [164]. We combined the estimates of the 30 different analyses (one for each imputed dataset, for each of which we followed all the previously highlighted frequentist two-stage meta-regression model approach) using Rubin's rules.

Results

Cohort and participant characteristics

The cohorts varied greatly in terms of geographic location, sample size, number of events and included a broad spectrum of patients with COPD from primary, secondary and tertiary care settings (Table 2). Mean FEV1 ranged from 30 to 70% of the predicted values, mean modified Medical Research Council (mMRC) dyspnea scores from 1.0 to 2.8 (the scale goes from 0 to 4, with 4 being the worst), mean number of exacerbations in the previous year (where available) from 0.2 to 1.7 and mean 6-minute walk distance (where available) from 218 to 487 meters.

Table 2 Study characteristics

Cohort	# Events	# patients	Person years	Mean age: years	Men: %	Mean FEV1% pred.	Mean mMRC	Past exacerbators: %	Mean # prev. exacer-bations	Current Smoker: %	Mean BMI	Mean 6MWT: meters	Mean SGRQ	Mean CAT
COPDgene	337	4484	10603	63 (9)	56	57.4 (22.8)	1.5	0.16		43	27.9 (6.1)	376.1 (124.1)	36.9 (22.9)	
Sevilla[a]	205	596	1562	66 (10)	95	43.5 (13.3)	1	0.25	1.16	24	29.2 (5.7)			
Copenhagen[b]	186	2287	6618	61 (9)	54	70.5 (23.7)	1.3			71	25 (4.2)			
Genkols	126	954	2708	65 (10)	61	46.9 (17)	1.3	0.15	0.6	47	25.4 (5)			
Zaragoza II[a]	118	1150	3069	63 (9)	93	62.3 (20.3)	1.1	0.17	0.91	34	27.5 (4.8)	356.2 (153.7)		
HUNT	116	1571	4583	63 (13)	62	63.8 (18.7)	1.3			47	26.4 (4.4)			
Galdakao[a]	92	543	1497	68 (8)	96	55 (13.3)	0.9	0	0.65	21	28.3 (4.4)	408.9 (92.4)		
Barmelweid[b]	79	232	555	72 (9)	60	45.2 (16.1)	1.1			21	26 (6.3)	363.4 (126.8)		
Terrassa III[a]	78	181	423	72 (10)	95	45.2 (14.4)	1.2	0.31	1.28	23	27.9 (5)	330.4 (105.8)		
Initiatives BPCO	76	930	1525	64 (10)	77	52.4 (20.3)	1.1	0.4	1.65	28	25.4 (5.5)	387.4 (120.8)	43.9 (19)	
Terrassa I[a]	72	135	284	72 (9)	92	41.3 (13)	1.3	0.25	1.03	17	26.3 (4.9)			
SEPOC[b]	61	318	871	65 (9)	100	45 (18.3)	1.5			38	26.4 (4.2)			
Requena II[a]	52	186	396	71 (9)	99	44.5 (16.5)	1	0.16	0.62	17	28.1 (5.2)	380.1 (111.9)		
ICE COLD ERIC	47	400	1071	67 (10)	57	55.3 (16.5)	1.5	0.13	0.58	39	26.1 (5.2)			
PAC-COPD[b]	41	342	980	68 (9)	93	52.4 (16.2)	1	0.04		33	28.2 (4.7)	435.5 (90.6)		
Tenerife[a]	34	275	653	63 (10)	79	55.8 (21.2)	1.2	0.06	0.37	42	27.3 (5.1)	487.4 (87.5)		
Terrassa II[a]	28	66	145	72 (9)	98	30.2 (12.9)	1	0.42	1.81	14	25.7 (4.3)	217.7 (76.6)		
Requena I[a]	23	174	393	72 (9)	99	48.1 (16.8)	1.2[c]	0.03	0.22	23	28 (4.2)	434.4 (125.3)		
Zaragoza I[a]	21	137	379	66 (8)	99	49.8 (17.6)	1.1			27	27.7 (4.6)	449 (91.9)		
Son Espases Mallorca	17	115	292	70 (7)	79	41.5 (13.4)	1	0.59		27	27.1 (5.9)	401.5 (89.7)		16.6 (8.2)

Basque[b]	16	106	299	71 (9)	98	46.9 (11.4)	0.6			23	26.1 (4.9)	442.9 (95.4)		
Japan	15	147	409	69 (7)	100	47.1 (17.5)	0.9			22	21 (2.9)		36.6 (16.5)	
La Princesa Madrid	11	318	633	71 (10)	74	50 (19.8)	1.1	0.18	0.77	19	26.2 (5.1)	337.1 (92.8)		
Pamplona[a]	7	190	470	65 (8)	84	68.9 (19.9)	1.1			37	27 (4.4)	463.2 (113.9)		
Mar de Plata Argentina	3	99	147	64 (9)	60	48.8 (18.6)	1	0.29		21	27 (5.6)	353.2 (128.7)		16.1 (7.8)
A1ATD[d]	0	308	834	58 (10)	60	53.1 (25.1)	1.2	0.52		5	25.7 (4.9)		50.8 (19.9)	20.5 (8.1)

Abbreviations: FEV1% pred.=forced expiratory volume in 1 second percentage predicted; mMRC=modified Medical Research Council (MMRC) dyspnea scale; past exacerbators are defined as patients with more than 1 exacerbation in the previous year; mean # previous exacerbations are referred to the previous year; BMI=body-mass index; 6MWT=6-minute walk test; SGRQ=Saint George's Respiratory Questionnaire; CAT=COPD Assessment Test.

The cohorts are presented in decreasing order of number of events. Most of the variables available provided by the 3CIA collaboration for the different cohorts are shown. In particular, we show all the variables constituting the scores analyzed in our study. We present the standard deviation for all individual variables, whose distribution is approximately normal; this is not the case for count (with small numbers) or categorical variables, like number of previous exacerbations or mMRC).

[a]Cohorts belonging to the COCOMICS collaboration

[b]Cohorts belonging to the ADO collaboration. For information concerning the cohorts, please see [40].

[c] Since none of the score could be evaluated in the cohort "Requena I" (mainly because the variable dyspnoea was missing for 95% of the patients, i.e. for 165 out of 174 patients), this cohort was excluded from the analysis.

[d] Since there was no event in a follow-up of 3 years, the cohort "A1ATD" was excluded from the analysis.

Missing cells correspond to variables that are completely missing in the cohort of the correspondent row.

Direct comparisons of prognostic scores for mortality in patients with COPD

The direct comparisons are shown in the upper-right triangle of Table 3, i.e. a league table (that also includes the Multiple Score Comparison Meta-Analysis in the lower-left triangle). Forest plots for each pair of scores, I^2 and τ^2 are shown in the appendix. 41 direct comparisons of the AUC of prognostic scores were possible; indeed, no direct evidence was available for the comparison between SAFE and the eBODE, BODEx, DOSE and B-AE-D scores (cells D6, E6, F2, G6, I10 in the league table 3).

Table 3 League table presenting the Multiple Score Comparison (MSC) meta-analysis (lower-left half of the table) and the direct random-effects meta-analysis (upper-right half of the table)

Direct meta-analysis MSC meta-analysis		GOLD	B-AE-D	GOLD (2011-2016)	DOSE	BODEx	SAFE	eBODE	BODE	BODE updated	ADO
		1	2	3	4	5	6	7	8	9	10
GOLD	A	AUC = 0.613 (1st Q. 0.587-3rd Q. 0.637)	ΔAUC = 0.030 (-0.005, 0.065)	0.017 (-0.005, 0.038)	0.036 (0.008, 0.064)	0.054 (0.029, 0.079)	0.047 (0.027, 0.068)	0.064 (0.004, 0.123)	0.071 (0.040, 0.102)	0.080 (0.041, 0.119)	0.090 (0.072, 0.109)
B-AE-D	B	ΔAUC = 0.010 (-0.010, 0.031)		-0.004 (-0.025, 0.017)	0.004 (-0.012, +0.020)	0.025 (0.011, 0.039)	NA	0.069 (-0.016, -0.121)	0.079 (0.004, 0.154)	0.082 (-0.012, -0.152)	0.076 (0.051, 0.101)
GOLD (2011-2016)	C	0.012 (-0.001, 0.024)	0.001 (-0.019, 0.021)		0.009 (-0.002, 0.021)	0.028 (0.017, 0.039)	0.055 (0.038, 0.072)	0.047 (0.016, 0.079)	0.059 (0.046, 0.073)	0.051 (0.027, 0.076)	0.067 (0.053, 0.080)
DOSE	D	0.022 (0.006, 0.037)	0.011 (-0.007, -0.029)	0.010 (-0.005, 0.025)		0.018 (0.008, -0.029)	NA	0.039 (0.007, 0.070)	0.033 (-0.000, 0.065)	0.043 (-0.002, 0.088)	0.061 (-0.044, -0.079)
BODEx	E	0.041 (0.027, 0.055)	0.030 (0.014, 0.047)	0.029 (0.015, 0.043)	0.019 (0.005, 0.033)		NA	0.030 (-0.001, 0.061)	0.031 (-0.017, 0.079)	0.039 (-0.028, 0.105)	0.050 (0.034, 0.066)
SAFE	F	0.061 (0.034, 0.087)	0.050 (0.018, 0.082)	0.049 (0.022, 0.076)	0.039 (0.010, 0.068)	0.020 (-0.009, 0.048)		NA	0.011 (-0.000, 0.023)	0.005 (-0.009, 0.018)	-0.007 (-0.029, 0.015)
eBODE	G	0.065 (0.046, 0.085)	0.055 (0.032, 0.078)	0.054 (0.034, 0.074)	0.044 (0.023, 0.064)	0.024 (0.007, 0.042)	0.005 (-0.025, 0.035)		-0.001 (-0.020, 0.017)	0.002 (-0.031, 0.034)	0.024 (-0.018, 0.066)
BODE	H	0.068 (0.052, 0.084)	0.057 (0.034, 0.080)	0.056 (0.039, 0.074)	0.046 (-0.027, -0.065)	0.027 (0.009, 0.045)	0.007 (-0.019, 0.034)	0.003 (-0.016, 0.021)		0.005 (-0.006, 0.017)	-0.004 (-0.023, 0.016)
BODE upd.	I	0.076 (0.058, 0.095)	0.066 (0.041, 0.091)	0.065 (0.045, 0.085)	0.055 (-0.033, -0.076)	0.036 (0.015, 0.056)	0.016 (-0.012, 0.043)	0.011 (-0.009, 0.031)	0.008 (-0.005, 0.022)		-0.005 (-0.032, 0.022)
ADO	L	0.083 (0.070, 0.096)	0.072 (0.052, 0.093)	0.071 (0.056, 0.087)	0.070 (-0.052, -0.089)	0.042 (0.026, 0.058)	0.022 (-0.005, 0.050)	0.018 (-0.003, 0.038)	0.015 (-0.002, 0.032)	0.007 (-0.012, 0.026)	

Abbreviations: AUC=Area Under the Curve; The lower-left half of the table refers to the Multiple Score Comparison meta-analysis. The upper-right half of the table refers to direct comparisons using conventional random effects meta-analysis. The first cell (first row, first column) gives a reference value (in red), namely the median and 1st and 3rd quartiles of the AUC of the GOLD classification across cohorts as an anchor to interpret the differences in AUC between the prognostic scores. In each other cell, each pair of scores is compared using the difference in AUC. For the lower-left half of the table we report in the correspondent cell the difference between the AUCs of the score in the row and the score in the column; instead, for the upper-right half of the table we report the difference between the AUCs of the score in the column and the score in the row or the. We decided for this representation to make a visual comparison between direct and MSC comparison easier; indeed, in this way, it is enough to look at corresponding values mirrored at the main diagonal. The 95% confidence interval is indicated in parentheses. For better readability of the table the sign "+" is omitted, while the sign "-" is indicated.

The updated BODE score performed statistically significantly better than GOLD, new GOLD and the B-AE-D scores whereas the AUC of the updated BODE score was higher than for the other scores but not statistically significantly so. We deemed overall statistical heterogeneity of direct comparisons moderate. In our MSC meta-analysis the direct comparisons should be interpreted with caution though, since they do not take into account that multiple scores were evaluated on the same set of patients and are thus likely to bias the interpretation of which prognostic score performs best [141].

Groups of cohorts evaluating the same prognostic scores

Grouping of cohorts where the same prognostic scores could be calculated was the first step to consider correlations introduced by predictions performed on the same sample of patients. Figure 1 shows the grouping of cohorts. In group 1 (constituted by 4 cohorts: Copenhagen, HUNT, Japan, SEPOC, as shown in Figure 1) information on FEV1, age and dyspnea was available to calculate the GOLD and ADO score for each participant. In contrast, group 6 consisted of 4 cohorts (La Princesa Madrid - Requena II - Tenerife - Terrassa II) where 9 prognostic scores (all except for the SAFE score) could be calculated for each participant. Figure 1 provides a visual representation of these groups together with the number of events (i.e. deaths). For example, the dark green line represents group 1 where the GOLD and ADO scores could be compared against each other. The closed polygons show the comparisons that are possible for each group of cohorts. Group 6 is represented by the dark yellow polygon that includes 9 scores. Thus, unlike multiple treatment network meta-analyses, where usually two or at most three treatments are compared in each trial, Figure 1 shows that in each of the cohorts of our database we can compare between 2 and 9 prognostic scores.

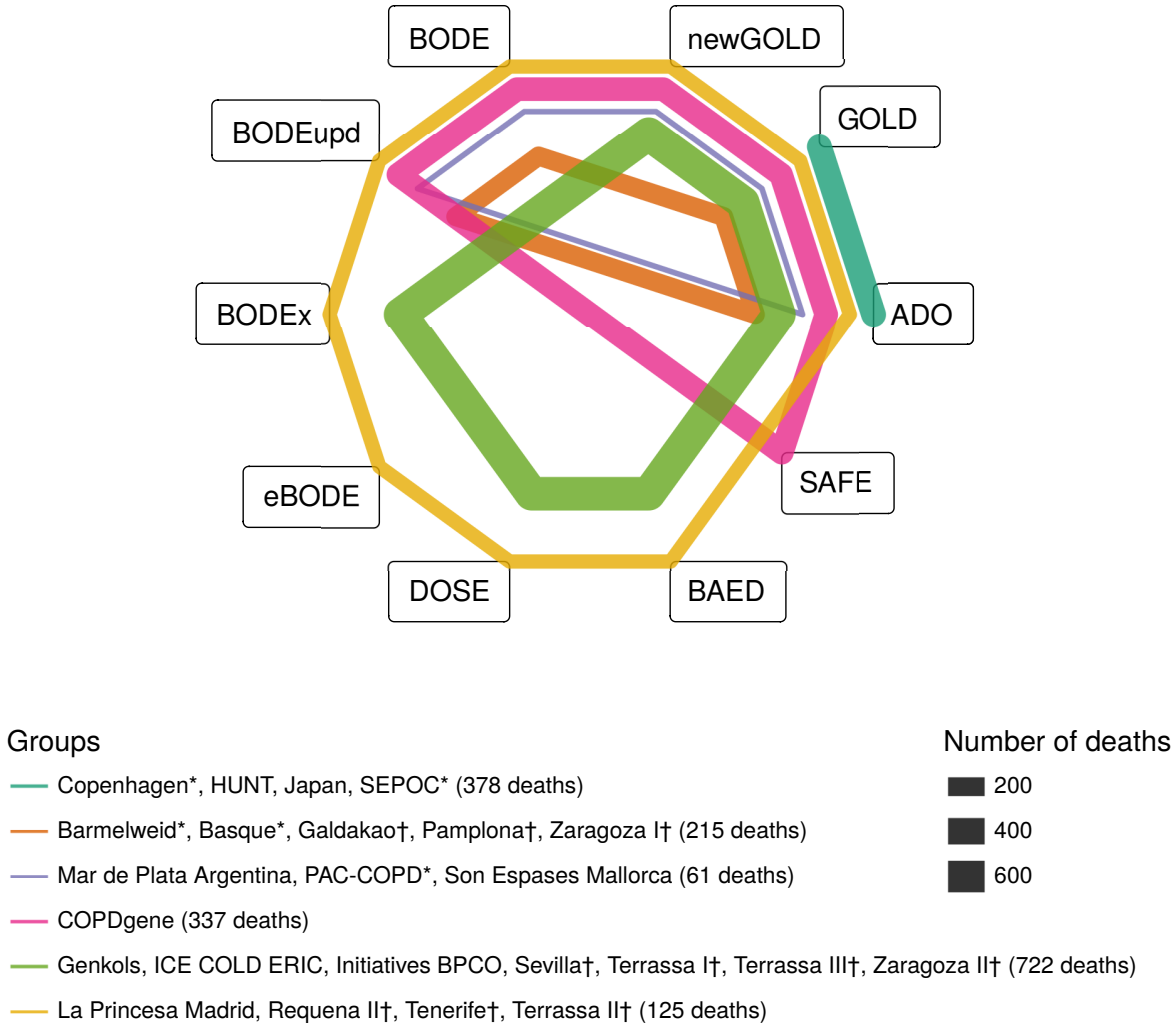


Figure 1 Network representing which prognostic scores belong to the different groups. Each node represents a score and each closed polygon represents a group of cohorts where the same prognostic scores are available. The thickness of the lines represents the total number of deaths in the specific group.

MSC meta-analysis of prognostic scores to predict 3-year mortality in patients with COPD

The lower-left part of Table 3 shows all comparisons between the AUCs of the 10 prognostic scores taking into account the correlation among multiple comparisons for the same patients as well as direct and indirect evidence of the entire network (Figure 1). The median AUC of the GOLD classification of airflow obstruction severity was 0.613 (interquartile range 0.587 to 0.637) and is shown in red in the upper-left cell as an anchor to interpret the differences in AUC between the prognostic scores.

Compared to GOLD, all prognostic scores showed statistically significantly higher AUCs except for the B-AE-D and GOLD 2011-2016 (cells B1-L1 in Table 3). Compared to the BODE score (the most commonly used prognostic score in COPD, median AUC 0.679 [interquartile range 0.655 to 0.733]), the ADO, updated BODE and eBODE showed higher AUCs whereas all other scores performed worse.

Figure 2 shows the comparisons of all scores against the BODE score and that the ADO score and the updated BODE performed better than the other scores (i.e., AUC_{ADO} - AUC_{BODE} =+0.015 [95% CI -0.002 to 0.032], p = 0.08; $AUC_{BODE\ updated}$ - AUC_{BODE} = 0.008 [95% CI = -0.005 to +0.022]; p = 0.23). The sensitivity analysis undertaken excluding from the database the 2 cohorts used in the large-scale update of the ADO index [55] shows no significant differences.

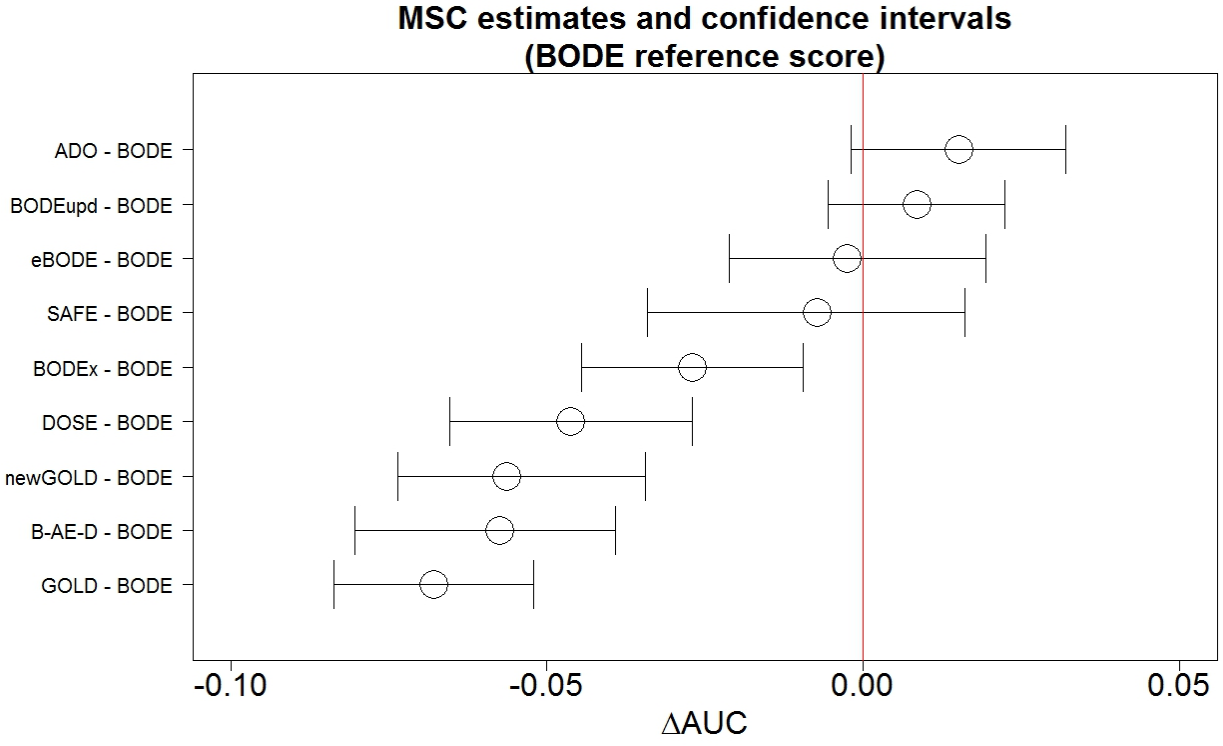


Fig. 2 Difference in AUC (shown with confidence interval with 95% confidence level) among the different scores and the BODE index (chosen here as the reference score) in the MSC meta-analysis. As a reference we use the median of the AUC of the BODE score 0.679 (1st Qu. 0.655, 3rd Qu. 0.733).

Heterogeneity, transitivity and Inconsistency

Global heterogeneity was relatively small (τ^2_{pooled} = 0.00011) (we did not use a τ^2 for each group (τ^2_g) since this is not recommended when there are groups with a single cohort [134]). The groups of the MSC meta-analysis were balanced with regard to characteristics of the different cohorts that may modify

the predictive performance of the scores (all a priori defined characteristics that were generating case-mix were not statistically significantly different across groups) and we could thus assume transitivity.

The consistency analyses did not suggest local or global inconsistency . Visual analysis of the Q-Q plot and studentized residuals indicated robust local consistency. The likelihood-ratio test statistic showed overall consistency (Q likelihood-ratio test = 25.29 $\cong \chi^2(0.95, 16) = 26.30$, p-value = 0.06).

Discussion

Our study has two main findings. Firstly, our results indicate that the ADO index has the best ability to predict 3-year mortality in patients with COPD, followed by the updated BODE and eBODE indices. Given its simplicity, the ADO index may be the most attractive option across care settings to inform patients and health care professionals about prognosis and to inform treatment decisions whose effectiveness may depend on life expectancy. Secondly, we presented a comprehensive approach for external validation and concurrent comparison of prognostic scores and its first application. MSC meta-analysis is a method adapted from network meta-analysis that meets the call for new approaches for external validation and concurrent comparison of risk prediction models and scores that should take advantage of data sharing, individual patient data (IPD) and advanced analytical techniques [31,33,34,43,182].

In practice, the GOLD score using just lung function is still used most commonly to grade disease severity, which is traditionally related to prognosis as in other fields (e.g. cancer). FEV1% pred. (thus, GOLD classification) is an important parameter at the population-level in the prediction of important clinical outcomes such as mortality and hospitalization. The revised combined COPD assessment and their further developments integrates the severity of airflow limitation assessment, providing also information regarding symptom burden and risk of exacerbation [50]. However, the results of our analysis show that, when the aims to predict mortality in individuals, other scores such as ADO, updated BODE and eBODE are substantially better than the GOLD classifications (in our analysis, GOLD and GOLD 2011-2016). We note that the AUC for the best score (ADO) is 0.69, a moderately good discriminative performance; however, we can often not expect a much higher discriminative performance in clinical settings (for instance, see [126]).

The predictive performance of a prognostic score is important, but it is not the only criterion to choose a prognostic score for practice. Indeed, with an eye towards applicability, the time, cost and burden for patients and practitioners to measure the predictors of a prognostic score should be taken into consideration [192]. We deem a prognostic score such as ADO to be easily available if it only includes simple questions, easily available information from medical charts and spirometry (performed for the diagnosis of COPD) [50,52].

Scores to predict mortality are also useful beyond estimating prognosis. Nowadays, no treatments to lower the risk of mortality are available for patients with COPD yet, thus for this outcome, prediction scores cannot provide risk-stratified treatment guidance. However, prognostic scores may help to make randomized trials with all-cause mortality as primary outcome more efficient than previous trials by only including patients at higher risks [101]. Also, prognostic scores for all-cause mortality are particularly attractive for multi-morbid patients such as COPD patients, where cardiovascular disease, diabetes, renal disease and lung cancer, among other conditions, also contribute to mortality [103,193]. Patients with COPD often receive less than optimal prevention and treatment of cardiovascular disease, which may partly reflect a therapeutic nihilism. Of course, there are patients who are unlikely to benefit

from longterm cardiovascular prevention because of short life expectancy. However, a prognostic score provides a better basis for decisions on cardiovascular prevention, lung cancer screening or other treatments and may limit under- and over-treatment in COPD [1,194,195].

Many prognostic models and scores (as the models' simplified form) are never validated in practice and many investigators develop a second model instead of relying on existing scores at least as a starting point. Such practice has led to numerous prognostic scores for the same conditions that are left without external validation. Thus, we introduced MSC meta-analysis, which addresses the lack of external validation and comparisons of prognostic scores by providing a comparison of their predictive performance in external validation cohorts and simultaneously considering the entire network of direct and indirect comparisons. Thereby, it allows for a comparison of predictive performance that is not limited by non-comparable spectrum of populations as it is commonly the case when evaluating the results of independent validation studies. MSC meta-analysis can be applied to any medical field, with the availability of individual patient data being the only major limiting factor.

Strengths of our study are the careful analytical approach to MSC meta-analysis and the availability of the R code that allow for widespread use and potential further development of the method. For the particular application of MSC meta-analysis here, a major strength is the large high-quality database of the 3CIA collaboration with the broadest possible COPD patient spectrum. The diverse case mix and broad patient spectrum greatly increases the probability that our results are generalizable to all COPD patients. A limitation of the study is that, ideally, a network meta-analysis is conducted prospectively and jointly planned for all of the cohorts involved to ensure equality of the clinical settings and homogeneity of study design, conduct and variable definitions. Though, this will rarely be the case in reality. Another limitation of our analysis is that we only used AUC as a performance measure, which we did for theoretical and practical reasons[180]. In general improvements in AUC have to be interpreted with caution [196]. Furthermore, we cannot exclude the possibility of case-mix effects due to variables that were not available in the database or unknown.

Further research needs include the extension of MSC to include measures of calibration, which is arguable as important as discrimination. For the area of COPD, it would be attractive to apply MSC to risk scores for exacerbations [50,197]. However, there are likely too few thoroughly developed and externally validated scores to predict exacerbations in patients with COPD [126]. Finally, given the large number of risk scores in the medical field and the lack of external validations and comparisons of risk scores, there is a great need for comparative studies that may use MSC in order to inform clinical practice and research about the most predictive scores [126].

Discussion

Our study has two main findings. Firstly, our results indicate that the ADO index has the best ability to predict 3-year mortality in patients with COPD, followed by the updated BODE and eBODE indices. Given its simplicity, the ADO index may be the most attractive option across care settings to inform patients and health care professionals about prognosis and to inform treatment decisions whose effectiveness may depend on life expectancy. Secondly, we presented a comprehensive approach for external validation and concurrent comparison of prognostic scores and its first application. MSC meta-analysis is a method adapted from network meta-analysis that meets the call for new approaches for

external validation and concurrent comparison of risk prediction models and scores that should take advantage of data sharing, individual patient data (IPD) and advanced analytical techniques [31,33,34,43,182].

In practice, the GOLD score using just lung function is used most commonly to grade disease severity, which is traditionally related to prognosis as in other fields (e.g. cancer). FEV1% pred. (thus, GOLD classification) is a very important parameter at the population-level in the prediction of important clinical outcomes such as mortality and hospitalization. The revised combined COPD assessment integrates the severity of airflow limitation assessment, providing also information regarding symptom burden and risk of exacerbation [50]. However, the results of our analysis show that, when the aims to predict mortality in individuals, other scores such as ADO, updated BODE and eBODE are substantially better than the GOLD classifications (in our analysis, GOLD and GOLD 2011-2016).

The predictive performance of a prognostic score is important, but it is not the only criterion to choose a prognostic score for practice. Indeed, with an eye towards applicability, the time, cost and burden for patients and practitioners to measure the predictors of a prognostic score should be taken into consideration [192]. We deem a prognostic score such as ADO to be easily available if it only includes simple questions, easily available information from medical charts and spirometry (performed for the diagnosis of COPD) [50,52]. We provide an example for how to balance the predictive performance and applicability of different prognostic scores in the supplementary material.

Scores to predict mortality are also useful beyond estimating prognosis. Nowadays, there are no treatments to lower the risk of mortality are available for patients with COPD yet, thus for this outcome, prediction scores cannot provide risk-stratified treatment guidance. However, prognostic scores may help to make randomized trials with all-cause mortality as primary outcome more efficient than previous trials by only including patients at higher risks [101]. Also, prognostic scores for all-cause mortality are particularly attractive for multi-morbid patients such as COPD patients, where cardiovascular disease, diabetes, renal disease and lung cancer also contribute to mortality [103]. Patients with COPD often receive less than optimal prevention and treatment of cardiovascular disease, which may partly reflect a therapeutic nihilism. Of course, there are patients who are unlikely to benefit from long term cardiovascular prevention because of short life expectancy. However, a prognostic score provides a better basis for decisions on cardiovascular prevention, lung cancer screening or other treatments and may limit under- and over-treatment in COPD [1,194,195].

Many prognostic models and scores (as the models' simplified form) are never validated in practice and many investigators develop a second model instead of relying on existing scores at least as a starting point. Such practice has led to numerous prognostic scores for the same conditions that are left without external validation. Thus, we introduced MSC meta-analysis, which addresses the lack of external validation and comparisons of prognostic scores by providing a comparison of their predictive performance in external validation cohorts and simultaneously considering the entire network of direct and indirect comparisons. Thereby, it allows for a comparison of predictive performance that is not limited by non-comparable spectrum of populations as it is commonly the case when evaluating the results of independent validation studies. MSC meta-analysis can be applied to any medical field, with the availability of individual patient data being the only major limiting factor.

Strengths of our study are the careful analytical approach to MSC meta-analysis and the availability of the R code that allow for widespread use and potential further development of the method. For the particular application of MSC meta-analysis here, a major strength is the large high-

quality database of the 3CIA collaboration with the broadest possible COPD patient spectrum. The diverse case mix and broad patient spectrum greatly increases the probability that our results are generalizable to all COPD patients. A limitation of the study is that, ideally, a network meta-analysis is conducted prospectively and jointly planned for all of the cohorts involved to ensure equality of the clinical settings and homogeneity of study design, conduct and variable definitions. Though, this will rarely be the case in reality. Another limitation of our analysis is that we only used AUC as a performance measure, which we did for theoretical and practical reasons (see supplementary material); in general improvements in AUC have to be interpreted with caution [196]. Furthermore, we cannot exclude the possibility of case-mix effects due to variables that were not available in the database or unknown.
Further research needs include the extension of MSC to include measures of calibration, which is arguable as important as discrimination. For the area of COPD, it would be attractive to apply MSC to risk scores for exacerbations [50,197]. However, there are likely too few thoroughly developed and externally validated scores to predict exacerbations in patients with COPD [126]. Finally, given the large number of risk scores in the medical field and the lack of external validations and comparisons of risk scores, there is a great need for comparative studies that may use MSC in order to inform clinical practice and research about the most predictive scores [126].

Conclusions

Borrowing from network meta-analysis, we presented a comprehensive approach for external validation and concurrent comparison of multiple prognostic scores. While our analyses showed best performance for the ADO and updated BODE scores to predict mortality for patients with COPD, MSC meta-analysis can be applied to prognostic scores in any the medical field to identify the best scores, possibly paving the way for stratified medicine, public health and research.

Additional Files

Additional file 1: Supplementary material (Word 90 kB)

List of abbreviations

COPD: chronic obstructive pulmonary disease; MSC: Multiple Score Comparison; NMA: network meta-analysis; GOLD: Global initiative for chronic Obstructive Lung Disease; BODE: Body mass index, airflow Obstruction, Dyspnea and severe Exacerbations; ADO: Age, Dyspnea, airflow Obstruction; e-BODE=severe acute exacerbation of COPD plus BODE; BODEx: Body mass index, airflow Obstruction, Dyspnea, severe acute Exacerbation of COPD; DOSE: Dyspnea, Obstruction, Smoking and Exacerbation frequency; SAFE: Saint George's Respiratory Questionnaire (SGRQ) score, Air-Flow limitation and Exercise capacity; B-AE-D: Body-mass index, Acute Exacerbations, Dyspnea; AUC: Area Under the Curve; SUCRA: surface under the cumulative ranking curve.

Declarations

Ethics approval and consent to participate

All cohorts were approved by ethics committees and participants gave written informed consent [40].

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are reported within the article elsewhere [40]. Progr amming language used for the main analysis: R version 3.0.2

Competing interests

The authors declare that they have not competing interests.

Funding

MAP obtained funding for the current study. All authors organized funding for their respective cohorts, which has been described in detail elsewhere [40].

Authors' Contributions

All the authors had full access to all of the data in the study and take responsibility for the integrity of the data. BG, SRH, MAP designed the study. All the authors contributed to the acquisition, analysis, or interpretation of data. BG, SRH and MAP drafted the manuscript. All authors provided a critical revision of the manuscript for important intellectual content. BG, SRH and MAP undertook the statistical analysis. All authors provided necessary support to contribute their data to the 3CIA collaboration. JBS and MAP supervised the study.

Acknowledgments

We would like to thank Sarah Crook, Violeta Gaveikaite, Laura Werlen and Alex Marzel (all University of Zurich, Zurich, Switzerland) for their comments.

References

1. Steyerberg EW. *Clinical Prediction Models*. (Gail M, Krickeberg K, Sarnet J, Tsiatis A, Wong W, eds.). Springer - Statistics for Biology and Health - ISBN: 978-1-4419-2648-7; 2010.
2. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: A meta-analysis. *Psychol Assess*. 2000;12(1):19-30. doi:10.1037//1040-3590.12.1.19.
3. Collins GS. The BMJ Gary Collins : Opening up multivariable prediction models. Bmj.
4. Van Gorp MJ, Steyerberg EW, Van Der Graaf Y. Decision Guidelines for Prophylactic Replacement of Bj??rk-Shiley Convexo-Concave Heart Valves: Impact on Clinical Practice. *Circulation*. 2004;109(17):2092-2096. doi:10.1161/01.CIR.0000125853.51637.C8.
5. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood

cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation*. 2014;129(25 SUPPL. 1). doi:10.1161/01.cir.0000437738.63853.7a.

6. World Health Organization. *Prevention of Cardiovascular Disease.*; 2007. doi:10.1001/jama.2011.1668.
7. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42(2):517-584. doi:10.1161/STR.0b013e3181fcb238.
8. Expert Panel on Detection, Evaluation and T of HBC in A. Executive summary of the third report of the national cholesterol education program Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Jama*. 2001;285(19):2486-2497. doi:10.1001/jama.285.19.2486.
9. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38. doi:10.1186/1472-6947-6-38.
10. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J Cancer*. 1982;45(3):361-366. doi:10.1038/bjc.1982.62.
11. Di M-Y, Liu H, Yang Z-Y, Bonis PAL, Tang J-L, Lau J. Prediction Models of Mortality in Acute Pancreatitis in Adults. *Ann Intern Med*. 2016;(38). doi:10.7326/M16-0650.
12. Herman WH. Predicting risk for diabetes: Choosing (or building) the right model. *Ann Intern Med*. 2009;150(11):812-814. doi:10.7326/0003-4819-150-11-200906020-00010.
13. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission A Systematic Review. *JAMA*. 2015;306(15).
14. Ma D, Watson L, Ja R, et al. Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes (Review) Risk scoring systems for predicting preterm birth with the aim of reducing associated adverse outcomes. 2011;(11). doi:10.1002/14651858.CD004902.pub4.Copyright.
15. Lamain-de Ruiter M, Kwee A, Naaktgeboren CA, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: prospective multicentre cohort study. *BMJ*. 2016;354(19):i4338. doi:10.1136/bmj.i4338.
16. Wilson PWF, D'Agostino RB, Levy D, Belanger a. M, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation*. 1998;97(18):1837-1847. doi:10.1161/01.CIR.97.18.1837.
17. Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. *Thorax*. 2014;69(7):616-622. doi:10.1136/thoraxjnl-2013-204155.
18. Justice AC, Covinsky KE, Berlin J a. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-524. doi:199903160-00009 [pii].
19. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353(i2416):11. doi:10.1136/bmj.i2416.
20. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? *Bmj*. 1995;311(7019):1539-1541. doi:10.1136/bmj.311.7019.1539.
21. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007.
22. Collins GS, de Groot J a, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40. doi:10.1186/1471-2288-14-40.
23. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research 1: what , why , and how ? *BMJ*. 2009;338(b375):1317-1320. doi:10.1136/bmj.b375.

24. Steyerberg EW, Moons KGM, Windt DA Van Der, et al. Prognosis Research Strategy (PROGRESS) 3 : Prognostic Model Research. *PLoS Med*. 2013;10(2 (e1001381)):9. doi:10.1371/jour-.
25. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247.
26. Bleeker S., Moll H., Steyerberg EW, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol*. 2003;56(9):826-832. doi:10.1016/S0895-4356(03)00207-5.
27. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research 3: validating a prognostic model. *BMJ*. 2009;338(b605):1432-1435. doi:10.1136/bmj.b605.
28. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453-473.
29. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586. doi:10.1002/sim.1844.
30. Vergouwe Y, Nieboer D, Oostenbrink R, et al. A closed testing procedure to select an appropriate method to update prediction models. *Stat Med*. 2014;Under revi(October). doi:10.1002/sim.7179.
31. Collins GS, Moons KGM. Comparing risk prediction models Should be routine when deriving a new model for the same purpose. *BMJ*. 2012;3186(May):1-2. doi:10.1136/bmj.e3186.
32. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):93-101. doi:10.1002/bjs.9723.
33. Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM. Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLoS Med*. 2015;12(10):1-12. doi:10.1371/journal.pmed.1001886.
34. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Assessing risk prediction models using individual participant data from multiple studies. *Am J Epidemiol*. 2014;179(5):621-632. doi:10.1093/aje/kwt298.
35. Debray TP a, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med 2013;32*. 2013;32:3158-3180. doi:10.1002/sim.5732.
36. Broeze KA, Opmeer BC, Bachmann LM, et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol*. 2009;9:22. doi:10.1186/1471-2288-9-22.
37. Ahmed I, Debray TP a, Moons KGM, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. *BMC Med Res Methodol*. 2014;14(1):3. doi:10.1186/1471-2288-14-3.
38. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. *Br Med J*. 2015;350(1088):1-9. doi:10.1136/bmj.h1088.
39. Siebeling L, ter Riet G, van der Wal WM, et al. ICE COLD ERIC--International collaborative effort on chronic obstructive lung disease: exacerbation risk index cohorts--study protocol for an international COPD cohort study. *BMC Pulm Med*. 2009;9:15. doi:10.1186/1471-2466-9-15.
40. Soriano JB, Lamprecht B, Ramírez AS, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. *Lancet Respir Med*. 2015;3(6):443-450. doi:10.1016/S2213-2600(15)00157-5.
41. Siebeling L, Puhan MA, Muggensturm P, Zoller M, Ter Riet G. Characteristics of Dutch and Swiss primary care COPD patients - baseline data of the ICE COLD ERIC study. *Clin Epidemiol*. 2011;3:273-283. doi:10.2147/CLEP.S24818.
42. Krumholz HM. Why data sharing should be the expected norm. *Bmj*. 2015;350(feb05 6):h599-

h599. doi:10.1136/bmj.h599.
43. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *Bmj*. 2016;353(i3140):11. doi:10.1136/bmj.i3140.
44. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *Bmj*. 2012;344(January):1-10. doi:10.1136/bmj.d7762.
45. Moons KGM, Schuit E. Prediction of cardiovascular disease worldwide. *Lancet Diabetes Endocrinol*. 2015;3(5):309-310. doi:10.1016/S2213-8587(15)00002-9.
46. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi:10.1016/j.jclinepi.2014.06.018.
47. Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: The COCOMICS study. *Eur Respir J*. 2013;42(2):323-332. doi:10.1183/09031936.00121012.
48. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol*. 1999;28(1):1-9. doi:10.1093/ije/28.1.1.
49. Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. *Heal (San Fr*. 2001;24(2):218-230.
50. Decramer M, Vogelmeier C, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obrstructive Pulmonary Disease (updated 2017). *Glob Initiat Chronic Obstr Lung Dis*. 2011.
51. Celli BR, MacNee W, Agusti A, et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J*. 2004;23(6):932-946. doi:10.1183/09031936.04.00014304.
52. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347-365. doi:10.1164/rccm.201204-0596PP.
53. Miravitlles M, Soler-Cataluña JJ, Calle M, et al. Spanish COPD Guidelines (GesEPOC): pharmacological treatment of stable COPD. Spanish Society of Pulmonology and Thoracic Surgery. *Arch Bronconeumol*. 2012;48(7):247-257. doi:10.1016/j.arbres.2012.04.001.
54. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(10):1005-1012. doi:10.1056/NEJMoa021322.
55. Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open*. 2012;2(6):1-10. doi:10.1136/bmjopen-2012-002152.
56. Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet*. 2009;374(9691):704-711. doi:10.1016/S0140-6736(09)61301-5.
57. Soler-Cataluña JJ, Martinez-Garcia MA, Sanchez LS, Tordera MP, Sanchez PR. Severe exacerbations and BODE index: two independent risk factors for death in male COPD patients. *Respir Med*. 2009;103(5):692-699. doi:http://dx.doi.org/10.1016/j.rmed.2008.12.005.
58. Esteban C, Quintana JM, Aburto M, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. *Respir Med*. 2011;105(11):1662-1670. doi:http://dx.doi.org/10.1016/j.rmed.2011.05.005.
59. Kostianev SS, Hodgev VA, Iluchev DH. Multidimensional system for assessment of COPD patients. Comparison with BODE index. *Folia Medica (Plovdiv)*. 2008;50(4):29-38.

60. Briggs A, Spencer M, Wang H, Mannino D, Sin DD. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. *Arch Intern Med*. 2008;168(1):71-79. doi:http://dx.doi.org/10.1001/archinternmed.2007.37.
61. Schembri S, Anderson W, Morant S, et al. A predictive model of hospitalisation and death from chronic obstructive pulmonary disease. *Respir Med*. 2009;103(10):1461-1467. doi:http://dx.doi.org/10.1016/j.rmed.2009.04.021.
62. Siebeling L, Musoro JZ, Geskus RB, et al. Prediction of COPD-specific health-related quality of life in primary care COPD patients: a prospective cohort study. *npj Prim Care Respir Med*. 2014;24(14060):7. doi:10.1038/npjpcrm.2014.60.
63. Esteban C, Quintana JM, Aburto M, Moraza J, Capelastegui A. A simple score for assessing stable chronic obstructive pulmonary disease. *QJM*. 2006;99(11):751-759. doi:10.1093/qjmed/hcl110.
64. Boeck L, Soriano JB, Brusse-Keizer M, et al. Prognostic assessment in COPD without lung function: The B-AE-D indices. *Eur Respir J*. 2016;47(6):1635-1644. doi:10.1183/13993003.01485-2015.
65. Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. *Chest*. 2005;127(6):1890-1897. doi:10.1378/chest.127.6.1890.
66. Cardoso F, Tufanin AT, Colucci M, Nascimento O, Jardim JR. Replacement of the 6-min walk test with maximal oxygen consumption in the BODE index applied to patients with COPD: An equivalency study. *Chest*. 2007;132(2):477-482. doi:10.1378/chest.07-0435.
67. Williams JEA, Green RH, Warrington V, Steiner MC, Morgan MDL, Singh SJ. Development of the i-BODE: Validation of the incremental shuttle walking test within the BODE index. *Respir Med*. 2012;106(3):390-396. doi:10.1016/j.rmed.2011.09.005.
68. Almagro P, Barreiro B, Ochoa de Echaguen A, et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. *Respiration*. 2006;73(3):311-317. doi:10.1159/000088092.
69. Thomsen M, Ingebrigtsen TS, Marott JL, et al. Inflammatory biomarkers and exacerbations in chronic obstructive pulmonary disease. *JAMA*. 2013;309(22):2353-2361. doi:http://dx.doi.org/10.1001/jama.2013.5732.
70. Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. *Am J Respir Crit Care Med*. 2009;180(12):1189-1195. doi:http://dx.doi.org/10.1164/rccm.200902-0271OC.
71. Müllerova H, Maselli DJ, Locantore N, et al. Hospitalized Exacerbations of COPD. *CHEST J*. 2015;147(4):999-1007. doi:10.1378/chest.14-0655.
72. Almagro P, Soriano JB, Cabrera FJ, et al. Comorbidities and Short-term Prognosis in Patients Hospitalized for Acute Exacerbation of COPD. *Chest*. 2012;142(5):1126-1133.
73. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2010;363(11):1128-1138. doi:10.1056/NEJMoa0909883.
74. Moberg M, Vestbo J, Martinez G, et al. Validation of the i-BODE index as a predictor of hospitalization and mortality in patients with COPD Participating in pulmonary rehabilitation. *Copd J Chronic Obstr Pulm Dis*. 2014;11(4):381-387. doi:http://dx.doi.org/10.3109/15412555.2013.836171.
75. Marin JM, Carrizo SJ, Casanova C, et al. Prediction of risk of COPD exacerbations by the BODE index. *Respir Med*. 2009;103(3):373-378. doi:http://dx.doi.org/10.1016/j.rmed.2008.10.004.
76. Ko FW, Tam W, Tung AH, et al. A longitudinal study of serial BODE indices in predicting mortality and readmissions for COPD. *Respir Med*. 2011;105(2):266-273. doi:http://dx.doi.org/10.1016/j.rmed.2010.06.022.
77. Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. *Am J Med Sci*. 2010;339(1):10-14. doi:http://dx.doi.org/10.1097/MAJ.0b013e3181bb8111.

78. Brusse-Keizer M, van der Palen J, van der Valk P, Hendrix R, Kerstjens H. Clinical predictors of exacerbation frequency in chronic obstructive pulmonary disease. *Clin Respir J*. 2011;5(4):227-234. doi:10.1111/j.1752-699X.2010.00234.x.
79. Takahashi T, Muro S, Tanabe N, et al. Relationship between periodontitis-related antibody and frequent exacerbations in chronic obstructive pulmonary disease. *PLoS One*. 2012;7(7):e40570. doi:10.1371/journal.pone.0040570.
80. Azarisman MS, Fauzi MA, Faizal MP, Azami Z, Roslina AM, Roslan H. The SAFE (SGRQ score, air-flow limitation and exercise tolerance) Index: a new composite score for the stratification of severity in chronic obstructive pulmonary disease. *Postgrad Med J*. 2007;83(981):492-497.
81. Motegi T, Jones RC, Ishii T, et al. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int J Copd*. 2013;8:259-271. doi:http://dx.doi.org/10.2147/COPD.S42769.
82. Gudmundsson G, Gislason T, Janson C, et al. Risk factors for rehospitalisation in COPD: role of health status, anxiety and depression. *Eur Respir J*. 2005;26(3):414-419. doi:10.1183/09031936.05.00078504.
83. Niewoehner DE, Lokhnygina Y, Rice K, et al. Risk indexes for exacerbations and hospitalizations due to COPD. *Chest*. 2007;131(1):20-28. doi:10.1378/chest.06-1316.
84. Jacob A, Laurin C, Boudreau M, Lavoie KL, Bacon SL. The impact of body mass index on chronic obstructive pulmonary disease exacerbations. *Can Respir J*. 2013;18(2):237-242.
85. Suetomo M, Kawayama T, Kinoshita T, et al. COPD assessment tests scores are associated with exacerbated chronic obstructive pulmonary disease in Japanese patients. *Respir Investig*. 2014;52:288-295.
86. Parshall MB, Mapel DW, Rice L, Williams A, O'Reilly J. Predictive validity of short-form health survey 36 items scales for chronic obstructive pulmonary disease exacerbation. *Hear Lung J Acute Crit Care*. 2008;37(5):356-365.
87. Bertens LC, Reitsma JB, Moons KGM, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Copd*. 2013;8:493-499. doi:http://dx.doi.org/10.2147/COPD.S49609.
88. Garcia-Aymerich J, Farrero E, Félez M a, Izquierdo J, Marrades RM, Antó JM. Risk factors of readmission to hospital for a COPD exacerbation: a prospective study. *Thorax*. 2003;58(2):100-105. doi:10.1136/thorax.58.2.100.
89. Chen Y. Readmission in Taiwan. *West J Nurs Res*. 2006;28(1):105-124.
90. Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2014;11(2):149-157. doi:http://dx.doi.org/10.1513/AnnalsATS.201307-243OC.
91. Omachi TA, Yelin EH, Katz PP, Blanc PD, Eisner MD. The COPD severity score: a dynamic prediction tool for health-care utilization. *Copd J Chronic Obstr Pulm Dis*. 2008;5(6):339-346. doi:http://dx.doi.org/10.1080/15412550802522700.
92. Lee SD, Huang MS, Kang J, et al. The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. *Respir Med*. 2014;108(4):600-608.
93. Amalakuhan B, Kiljanek L, Parvathaneni A, Hester M, Cheriyath P, Fischman D. A prediction model for COPD readmissions: catching up, catching our breath, and improving a national problem. *J Community Hosp Intern Med Perspect*. 2012;2(1):9915. doi:http://dx.doi.org/10.3402/jchimp.v2i1.9915.
94. Ong KC, Earnest A, Lu SJ. A multidimensional grading system (BODE index) as predictor of hospitalization for COPD. *Chest*. 2005;128(6):3810-3816.
95. Echave-Sustaeta J, Comeche Casanova L, Garcia Lujan R, Sayas Catalan J, Gomez de la Camara A, Lopez Encuentra A. Prognosis following acute exacerbation of COPD treated with non-invasive

mechanical ventilation. *Arch Bronconeumol*. 2010;46(8):405-410. doi:http://dx.doi.org/10.1016/j.arbres.2010.03.004.
96. Decramer M. Management of COPD according to guidelines. A national survey among Belgian physicians. *Monaldi Arch Chest Dis*. 2003.
97. Rutschmann OT. Knowledge of guidelines for the management of COPD: a survey of primary care physicians. *Respir Med*. 2004.
98. Puhan MA, Zoller M, Ter Riet G. COPD : more than respiratory (comment). *Lancet (comment)*. 2008;371:26-27.
99. Fritsch K, Jacot M-L, Klarer A, et al. Adherence to the Swiss guidelines for management of COPD: experience of a Swiss teaching hospital. *Swiss Med Wkly*. 2005;135(7-8):116-121. doi:2005/07/smw-10844.
100. Calverley PMA, Anderson JA, Celli B, et al. Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2007;356(8):775-789.
101. Vestbo J, Anderson JA, Brook RD, et al. Fluticasone furoate and vilanterol and survival in chronic obstructive pulmonary disease with heightened cardiovascular risk (SUMMIT): A double-blind randomised controlled trial. *Lancet*. 2016;387(10030):1817-1826. doi:10.1016/S0140-6736(16)30069-1.
102. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet*. 2007;370(9589):797-799. doi:10.1016/S0140-6736(07)61383-X.
103. Fabbri LM, Luppi F, Beghé B, Rabe KF. Complex chronic comorbidities of COPD. *Eur Respir J*. 2008;31(1):204-212. doi:10.1183/09031936.00114307.
104. Hamburg MA, Collins FS. The Path to Personalized Medicine. *N Engl J Med*. 2010;363(4):301-304. doi:10.1056/NEJMp1002530.
105. Weston AD, Hood L. Systems Biology , Proteomics , and the Future of Health Care : Toward Predictive , Preventative , and Personalized Medicine Introduction : Paradigm Changes in Health Care. *J Proteome Res*. 2004;3(2):179-196. doi:10.1021/pr0499693.
106. Agusti A. The path to personalised medicine in COPD. *Thorax*. 2014;0(4):1-8. doi:10.1056/NEJMp1002530.
107. Agusti A, Sobradillo P, Celli BR. Addressing the complexity of chronic obstructive pulmonary disease: from phenotypes and biomarkers to scale-free networks, systems biology, and P4 medicine. *Am J Respir Crit Care Med*. 2011;183(9):1129-1137. doi:10.1164/rccm.201009-1414PP.
108. Decramer M, Vestbo J. *Global Initiative for Chronic Obstructive Lung Disease (Updated 2014)*.; 2014.
109. Puhan MA, Bachmann LM, Kleijnen J, Ter Riet G, Kessels AG. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Med*. 2009;7:2. doi:10.1186/1741-7015-7-2.
110. Hingorani ADA, Windt DVD a Van Der, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj*. 2013;346(February 2013):e5793. doi:10.1136/bmj.e5793.
111. Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. *BMC Med*. 2013;11(1):7. doi:10.1186/1741-7015-11-7.
112. *Systematic Reviews (CRD Guidance)*. Center for Reviews and Dissemination, University of York; 2009.
113. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *Annu Intern Med*. 2009;151(4):264-269. doi:10.1371/journal.pmed1000097.
114. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and

elaboration. *PLoS Med*. 2009;6(7):e1000100. doi:10.1371/journal.pmed.1000100.
115. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55-64. doi:10.7326/M14-0697.
116. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.
117. Cochrane Statistical Methods group and Cochrane Bias Methods Group. *Assessing Risk of Bias in Included Studies*.; 2011.
118. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-781.
119. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690. doi:10.1136/heartjnl-2011-301246.
120. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research 4: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(7709):1487-1490. doi:10.1136/bmj.b606.
121. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61(11):1085-1094. doi:10.1016/j.jclinepi.2008.04.008.
122. Alonzo TA. Clinical Prediction Models: A Practical Approach to Development, Validation , and Updating (Book Review). *AJE (American J Epidemiol*. 2009;170(4):90033. doi:10.1093/aje/kwp129.
123. Khalili D, Hadaegh F, Soori H, Steyerberg EW, Bozorgmanesh M, Azizi F. Clinical Usefulness of the Framingham Cardiovascular Risk Profile Beyond Its Statistical Performance The Tehran Lipid and Glucose Study. *AJE (American J Epidemiol*. 2012;176(3):177-186. doi:10.1093/aje/kws204.
124. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Risk Factor Models for Neurodevelopmental Outcomes in Children Born Very Preterm or With Very Low Birth Weight : A Systematic Review of Methodology and Reporting. *AJE (American J Epidemiol*. 2017;185(7):601-612. doi:10.1093/aje/kww135.
125. Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF. Predicting Risk of Type 2 Diabetes Mellitus with Genetic Risk Models on the Basis of Established Genome-wide Association Markers : A Systematic Review. *AJE (American J Epidemiol*. 2013;178(8):1197-1207. doi:10.1093/aje/kwt123.
126. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev*. 2017;26(143):1-13. doi:10.1183/16000617.0061-2016.
127. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *AJE (American J Epidemiol*. 2006;163(9):783-789. doi:10.1093/aje/kwj093.
128. Collins GS, Moons KGM. Comparing risk prediction models. *Br Med J*. 2012;344(may24 2):e3186-e3186. doi:10.1136/bmj.e3186.
129. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733-2749. doi:10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.
130. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(November 2001):2313-2324. doi:10.1002/sim.1201.
131. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105-3124. doi:10.1002/sim.1875.
132. Lu G, Ades a. E. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *J Am Stat*

Assoc. 2006;101(474):447-459. doi:10.1198/016214505000001302.
133. Salanti G, Higgins J, Ades a E, Ioannidis JP a. Evaluation of networks of randomized trials. *Stat Methods Med Res*. 2008;17(3):279-301. doi:10.1177/0962280207080643.
134. Lu G, Welton NJ, Higgins J, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Res Synth Methods*. 2011;2(1):43-60. doi:10.1002/jrsm.34.
135. Zarin W, Veroniki AA, Nincic V, et al. Erratum to: Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med*. 2017;15(1):61. doi:10.1186/s12916-017-0832-6.
136. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385.
137. Panagiotou OA. Network Meta-Analysis : Evidence Synthesis With Mixed Treatment Comparison (Book Review). *AJE (American J Epidemiol*. 2015;181(4):288-289. doi:10.1093/aje/kwu471.
138. Takwoingi Y. Empirical Evidence of the Importance of Comparative Studies of Diagnostic Test Accuracy. *Ann Intern Med*. 2013;158.
139. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res*. 2015:0962280215592269-. doi:10.1177/0962280215592269.
140. Manolio TA, Weis BK, Cowie CC, et al. New Models for Large Prospective Studies : Is There a Better Way ? *AJE (American J Epidemiol*. 2012;175(June 2010):859-866. doi:10.1093/aje/kwr453.
141. Franchini a. J, Dias S, Ades a. E, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. *Res Synth Methods*. 2012;3(2):142-160. doi:10.1002/jrsm.1049.
142. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(March):80-97. doi:10.1002/jrsm.1037.
143. Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology*. 1983;148:839-843.
144. Borenstein M, Hedges LV., Higgins J, R. RH. *Introduction to Meta-Analysis*. Wiley; 2011.
145. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2.
146. Nikolakopoulou A, Mavridis D, Salanti G. How to interpret meta-analysis models: fixed effect and random effects meta-analyses. *Evid Based Ment Health*. 2014;17(2).
147. Nikolakopoulou A, Mavridis D, Salanti G. Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health*. 2014;17(2):53-57.
148. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*. 2012;12(1):152. doi:10.1186/1471-2288-12-152.
149. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods*. 2013;4(4):291-323. doi:10.1002/jrsm.1085.
150. Cooper NJ, Sutton AJ, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*. 2009;28(15):1982-1998. doi:10.1002/sim.
151. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med*. 2013;11(1):159. doi:10.1186/1741-7015-11-159.

152. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and Technical Challenges in Network Meta-analysis. *Ann Intern Med*. 2013;159:130-137. doi:10.7326/0003-4819-159-2-201307160-00008.
153. Efthimiou O, Debray TP a, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;(November 2014). doi:10.1002/jrsm.1195.
154. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins J. Evaluating the quality of evidence from a network meta-analysis. *PLoS One*. 2014;9(7):e99682. doi:10.1371/journal.pone.0099682.
155. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 2010;172(8):971-980. doi:10.1093/aje/kwq223.
156. Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-1573. doi:10.1002/sim.1187.
157. Dias S, Welton NJ, Caldwell DM, Ades a. E. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(November 2008):932-944. doi:10.1002/sim.3767.
158. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley Serie in Probability and Mathematical Statistics; 1987. doi:10.1002/9780470316696.
159. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Med Res Methodol*. 2010;10(1):112. doi:10.1186/1471-2288-10-112.
160. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for Misclassified Outcomes in Binary Regression Models Using Multiple Imputation With Internal Validation Data. *AJE (American J Epidemiol*. 2013;177(9):904-912. doi:10.1093/aje/kws340.
161. Groenwold RHH, Donders ART, Roes KCB, Harrell FE, Moons KGM. Dealing With Missing Outcome Data in Randomized Trials and Observational Studies (Practice of Epidemiology). *AJE (American J Epidemiol*. 2012;175(3):210-217. doi:10.1093/aje/kwr302.
162. White IR. Network meta-analysis. *Stata J*. 2015;15(4):951-985. doi:The Stata Journal.
163. Pepe MS. Problems With Risk Reclassification Methods for Evaluating Prediction Models. *AJE (American J Epidemiol*. 2011;173(11):1327-1335. doi:10.1093/aje/kwr013.
164. Harrell FE. *Regression Modelling Strategies*. (Bickel P, Diggle P, Feinberg SE, Gather U, Olkin I, Zeger S, eds.). Springer - Statistics for Biology and Health; ISBN: 978-3-319-19424-0; 2015.
165. Laupacis A, Wells G, Richardson W. S, Tugwell P. Users' guides to the medical literature: V. How to use an article about prognosis. *Jama*. 1994;272(3):234-237. doi:doi:10.1001/jama.1994.03520030076032.
166. Puhan MA, Yu T, Stegeman I, Varadhan R, Singh S, Boyd CM. Benefit-harm analysis and charts for individualized and preference-sensitive prevention: example of low dose aspirin for primary prevention of cardiovascular disease and cancer. *BMC Med*. 2015;13(250):11. doi:10.1186/s12916-015-0493-2.
167. Hayes DF, Markus HS, Leslie RD, Topol EJ. Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC Med*. 2014;12(37):8. doi:10.1186/1741-7015-12-37.
168. Nickel CH, Bingisser R, Morgenthaler NG. The role of copeptin as a diagnostic and prognostic biomarker for risk stratification in the emergency department. *BMC Med*. 2012;10(7). doi:10.1186/1741-7015-10-7.
169. Vuong K, McGeechan K, Armstrong BK, Cust AE. Risk Prediction Models for Incident Primary Cutaneous Melanoma: A Systematic Review. *JAMA Dermatology*. 2014;150(4):434. doi:10.1001/jamadermatol.2013.8890.
170. Kagen D, Theobald C, Freeman M. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA - J Am Med Assoc*. 2011;306(15):1688-1698.
171. Yourman LC, Lee SJ, Schonberg M a, Widera EW, Smith AK. Prognostic Indices for Older Adults: A

Systematic Review. *JAMA*. 2012;307(2):182-192.
172. Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9(103):14. doi:10.1186/1741-7015-9-103.
173. Esteban C, Quintana JM, Moraza J, et al. BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice? *BMC Med*. 2010;8:28. doi:10.1186/1741-7015-8-28.
174. Quintana JM, Esteban C, Unzurrunzaga A, et al. Predictive score for mortality in patients with COPD exacerbations attending hospital emergency departments. *BMC Med*. 2014;12:66. doi:http://dx.doi.org/10.1186/1741-7015-12-66.
175. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research 2: Developing a prognostic model. *BMJ (Research Methods Reporting)*. 2009;338(june):1373-1377. doi:10.1136/bmj.b604.
176. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med*. 2010;8(21):11. doi:1741-7015-8-21 [pii] 10.1186/1741-7015-8-21.
177. Hemingway H. Ten steps towards improving prognosis research Problems with prognosis research. *BMJ*. 2014;4184(December 2009):1-10.
178. Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC Med*. 2016;14(104):15. doi:10.1186/s12916-016-0650-2.
179. Thangaratinam S, Allotey J, Marlin N, et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med*. 2017;15(68):11. doi:10.1186/s12916-017-0827-3.
180. Haile SR, Guerra B, Soriano JB, Puhan MA. Multiple Score Comparison: A network meta-analysis approach to comparison and external validation of prognostic scores. *Stat Methods Med Res*. 2017;17(172):1-12. doi:(submitted).
181. Kessels AG, Riet G, Puhan MA, Kleijnen J, Bachmann LM, Minder C. A simple regression model for network meta-analysis. *OA Epidemiol*. 2013;1(1):1-8.
182. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med*. 2011;9(79):5. doi:10.1186/1741-7015-9-79.
183. Salanti G, Marinho V, Higgins J. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol*. 2009;62(8):857-864. doi:10.1016/j.jclinepi.2008.10.001.
184. Sauter R, Held L. Network meta-analysis with integrated nested Laplace approximations. *Biometrical J*. 2015;57(6):1038-1050. doi:10.1002/bimj.201400163.
185. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating (ROC) Curvel Characteristic. *Radiology*. 1982;143:29-36.
186. Caldwell DM, Ades a E, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj*. 2005;331(7521):897-900. doi:10.1136/bmj.331.7521.897.
187. Mills EJ, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. *Jama*. 2012;308(12):1246-1253. doi:10.1001/2012.jama.11228.
188. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *Bmj*. 2017;356(i6460):11. doi:10.1136/bmj.i6460.
189. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: An overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-171. doi:10.1016/j.jclinepi.2010.03.016.

190. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol*. 2015;15(58):9. doi:10.1186/s12874-015-0060-8.
191. Ransohoff DF, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-930.
192. Ioannidis JP a, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. *JAMA - J Am Med Assoc*. 2010;303(16):1646-1647. doi:10.1001/jama.2010.503.
193. Divo M, Cote C, De Torres JP, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155-161. doi:10.1164/rccm.201201-0034OC.
194. Reilly BM, Evans AT, Schaider JJ, et al. Impact of a Clinical Decision Rule in the Emergency Department. *JAMA - J Am Med Assoc*. 2002;288(3):342-350.
195. McGinn TG, Guyatt GH, Wyer PC, David Naylor C, Stiell IG, Scott Richardson W. Users' Guides to the Medical Literature. *J Am Med Assoc*. 2000;284(1):79-84. doi:10.1001/jama.284.1.79.
196. Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA - J Am Med Assoc*. 2009;302(21):2345-2352. doi:10.1001/jama.2009.1757.
197. Wedzicha JA, Brill SE, Allinson JP, Donaldson GC. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. *BMC Med*. 2013;11(181):10. doi:10.1186/1741-7015-11-181.
198. Parkes G. Using risk scores with patients. *Bmj*. 2010;340(jun22 3):c3327-c3327. doi:10.1136/bmj.c3327.
199. Steyerberg EW, Bleeker S., Moll H a, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441-447. doi:10.1016/S0895-4356(03)00047-7.
200. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005.
201. Reilly BM, Evans AT. Academia and clinic. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.
202. Held U, Garcia-Aymerich J, Ter Riet G, Moons KGM, Puhan MA. Missing variables in risk prediction models. *AJE (American J Epidemiol*. 2012. doi:10.1002/elan.
203. Harrell FE. *Regression Modeling Strategies*. Springer; 2015.
204. Snell KIE, Hua H, Debray TPA, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2016;69:40-50. doi:10.1016/j.jclinepi.2015.05.009.
205. Blattenberger G, Lad F. Separating the Brier Score into Calibration and Refinement Components : A Graphical Exposition Author (s): Gail Blattenberger and Frank Lad Source : The American Statistician , Vol . 39 , No . 1 (Feb ., 1985), pp . 26-32 Published by : Taylor & Franci. *Am Stat*. 1985;39(1):26-32.
206. Murphy AH. A New Vector Partition of the Probability Score. *J Appl Metereology*. 1973;12:595-600.

CHAPTER 5: General Discussion

State of the art

This thesis has focused on prediction models with a practical application to COPD. Prediction models are supposed to help in clinical practice, public health management and medical research. However, in most clinical fields, the path from model development to clinical application still has a long way to go.[1] Some medical fields are already in the stage of clinical application, being supported by guidelines, which strongly recommend the use of prediction models. For instance, in cardiovascular disease, the famous Framingham risk score is already widely accepted[2–5] and is used in clinical practice to complement the assessment of the clinicians; for instance, in the UK general practitioners have routine access to risk score calculators.[6] In other fields, like COPD, the use of prediction models is still half way through; some promising models were developed, but they are mostly used in research but not widely accepted by the medical community yet. Clinicians are often reluctant to use models that do not look intuitive, neither doctor- nor patient-friendly, which are thus not easy to use in practice.[6] In particular, one of the most famous COPD guidelines reports only a brief paragraph to speak about only one of the dozens of models developed in the field.[7]

The skepticism of clinicians is in part justified, because some prerequisites for the use of prediction models are not met. In particular, for a model to be reliable for clinical practice, it has to satisfy to following requirements:

- To inform specific decisions.
- To follow a sound methodological development.[8,9]
- To show accuracy (i.e., prognostic properties in the derivation cohort in which it was developed)
- To show good reproducibility (i.e., a good performance of the model in the same source population of the derivation cohort, for example by bootstrapping or cross-validation)[8–13]
- To show good transportability (i.e., a good performance of the model in a different population or "out-of-sample", by external validation).[10,14–18]
- To be of proven clinical effectiveness (i.e. as shown by an impact study).[18–20]

This dissertation has addressed some of these critical points both from a methodological and COPD-specific perspective and provides a basis from moving forward the use of prediction model in practice and in research.

The methodological contribution

We took up the challenge (highlighted several times in top journals)[21,22] for new approaches of external validation and concurrent comparison of risk prediction models that should take advantage of "big data" sharing,[23,24] individual patient data (IPD) and advanced analytical techniques.[22,25–31] Indeed, widespread recognition of the value of collaborative initiatives, together with medical registries (where either active or passive follow-up procedures are in place to capture disease occurrence or outcomes) and

implementation of medical information systems (that have similar data capture capability as registries) all offer great opportunities for data sharing and exploitation which counter this problem.

This thesis provides the new methodology of MSC meta-analysis that allows for extensive external validation of all prediction models that are assessed in a given number of cohort studies. Until now, prediction models were, if at all, validated in one or few external cohort studies. MSC meta-analysis probably provides the most definitive estimates of prediction model performance possible since the estimates are based on as many cohort studies that are available. We had the privilege of having 26 cohort studies of the COPD Cohorts Collaborative International Assessment (3CIA) consortium[32] available to apply MSC meta-analysis for the first time. In addition, MSC meta-analysis allows for the comparison of prediction models, which is rarely done although of great importance for those who want to use the most accurate and applicable prognostic model. MSC meta-analysis can be applied in any medical field and contribute to a better evidence base on the (comparative) performance of prediction models.

The contribution to the area of COPD

This thesis provides valuable information on prediction models for exacerbations and mortality. The systematic review clearly showed that none of the prediction models for exacerbations are ready for use yet. Most of the 27 prediction models were not developed following rigorous methodological standards and almost no one of them has been validated. Even if one wanted to compare the performance of these prediction models great caution is needed because exacerbations were defined and measured in different ways and because the statistical models were so diverse that they serve different aims for predicting exacerbations. This is a pity since the prediction of exacerbations would be much needed for risk-stratified treatment and to ultimately improve COPD care by reducing the risk for exacerbations that have a detrimental effect on health-related quality of life, survival and health-care cost.

In contrast to exacerbations, the evidence base on the performance of prediction models for death is now strong, thanks to the first application of MSC meta-analysis to a large number of COPD patients from diverse settings. There is clearly room for improving their predictive performance, but for the existing prognostic scores evaluated in this thesis, there is now strong evidence on their (comparative) performance. Our MSC meta-analysis showed that the ADO and BODE updated index clearly perform best and the easy availability of the three predictors of the ADO index speaks for adopting this score in research and practice.[33] While risk-stratified treatment of COPD based on mortality may not be at the core of COPD management, the prediction of mortality may guide clinicians in the use of other therapies (e.g. for cardiovascular disease, which is highly prevalent in COPD patients) and help reducing some of the nihilistic approach towards treating COPD patients beyond COPD-specific therapies. Also, it may help to inform the decision for and against lung cancer screening, which is a relevant issue for COPD patients, who are at substantially higher risk of lung cancer than people without COPD: some guidance currently suggests not to screen COPD patients because of their high risk for complications if lung cancer is detected and treated. However, the clinical manifestations of COPD are highly heterogeneous as is their prognosis so that a risk-stratified approach to screening, which is recommended anyway (i.e. guided by smoking history and other factors), based also on mortality may be attractive.

Finally, the ADO and updated BODE scores as the most accurate scores can be used to identify COPD patients for trials, that have specific eligibility criteria in terms of mortality. Some trials may want to identify patients at low risk for mortality, whereas other trials with mortality as an outcome may go for moderate to high risk patients in order to make these trials as efficient as possible in terms of sample sizes needed. The prognostic scores can also be used for control of confounding, either through pre-stratification in trials or through statistical adjustment of analysis of trials or cohort studies.

What is still missing?

Starting from this thesis, possible future research projects contributing to the applicability of prediction models are highlighted in the following paragraphs.

Comparison with a competitor approach

As we have seen in details previously, the MSC meta-analysis approach relies on a group structure that is defined by the patterns of missing data (in the same group we have cohorts with the same scores available, according to the variables collected for each database). As opposed to the MSC meta-analysis, a second promising approach for external validation and comparison of prediction models is to impute all the missing data, not only for the variables existing in the cohort databases, but also for variables completely missing in individual databases. Imputing a whole variable in a cohort could sound not intuitive or not reliable; however, recent studies support the imputation or predictors even in the extreme scenario of variable completely missing.[34] With this approach, the patterns of missing data would disappear and we would not need a network meta-analysis approach anymore. At first glance, the MSC meta-analysis should be a more conservative approach when several variables are completely missing in several cohorts; instead, when a small fraction of variables are missing in a small fraction of cohorts, the "whole-variable imputation" technique could be more precise and less computationally burdensome.[8,35] A need for medical research is to further explore and compare these two methodologies and to understand in which case the one outperforms the other, in terms of precision and reliability, according to the available database (in particular the patterns of missing data).

A comprehensive assessment of the performance of prediction models

It is of great importance to extend the MSC methodology to calibration properties.[36,37] Indeed, MSC was initially developed to evaluate the discriminative performance. It needs to be extended to calibration properties (discrimination and calibration properties being the ones in which the overall performance measures of prediction models can be decomposed; for instance, this can be formally done for the Brier score).[38,39] One possible candidate as a calibration performance measure is the "calibration slope".[36]

Concluding remarks

We shed light on the current state of models predicting exacerbations in patients with COPD, highlighting that none of the twenty-seven prediction models for COPD exacerbations appears to be ready to support

personalized COPD treatment or research. Furthermore, we filled a main gap in the area of prediction models by developing the methods for MSC meta-analysis. Thereby, large-scale external validation and comparison of prediction models become possible. The application of MSC meta-analysis suggests that the ADO index may be the best suited score to use in clinical practice for prediction of mortality in patients with COPD since it performed best among all scores and since its components are easily available.

MSC meta-analysis can be applied to any field of medicine and addresses the great need for external validation and comparison of prediction models. Thereby, the best prediction models can be identified paving the way for risk-stratified, personalized medicine.

References

1. Steyerberg EW, Moons KGM, Windt DA Van Der, et al. Prognosis Research Strategy (PROGRESS) 3 : Prognostic Model Research. *PLoS Med*. 2013;10(2 (e1001381)):9. doi:10.1371/jour-.
2. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation*. 2014;129(25 SUPPL. 1). doi:10.1161/01.cir.0000437738.63853.7a.
3. World Health Organization. *Prevention of Cardiovascular Disease.*; 2007. doi:10.1001/jama.2011.1668.
4. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42(2):517-584. doi:10.1161/STR.0b013e3181fcb238.
5. Expert Panel on Detection, Evaluation and T of HBC in A. Executive summary of the third report of the national cholesterol education program Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Jama*. 2001;285(19):2486-2497. doi:10.1001/jama.285.19.2486.
6. Parkes G. Using risk scores with patients. *Bmj*. 2010;340(jun22 3):c3327-c3327. doi:10.1136/bmj.c3327.
7. Decramer M, Vogelmeier C, Agustí AG, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obrstructive Pulmonary Disease (updated 2017). *Glob Initiat Chronic Obstr Lung Dis*. 2011.
8. Steyerberg EW. *Clinical Prediction Models*. (Gail M, Krickeberg K, Sarnet J, Tsiatis A, Wong W, eds.). Springer - Statistics for Biology and Health - ISBN: 978-1-4419-2648-7; 2010.
9. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690. doi:10.1136/heartjnl-2011-301246.
10. Steyerberg EW, Bleeker S., Moll H a, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56(5):441-447. doi:10.1016/S0895-4356(03)00047-7.
11. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and

external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005.
12. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774-781.
13. Harrell FE. *Regression Modelling Strategies*. (Bickel P, Diggle P, Feinberg SE, Gather U, Olkin I, Zeger S, eds.). Springer - Statistics for Biology and Health; ISBN: 978-3-319-19424-0; 2015.
14. Justice AC, Covinsky KE, Berlin J a. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-524. doi:199903160-00009 [pii].
15. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research 3: validating a prognostic model. *BMJ*. 2009;338(b605):1432-1435. doi:10.1136/bmj.b605.
16. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-2586. doi:10.1002/sim.1844.
17. Bleeker S., Moll H., Steyerberg EW, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol*. 2003;56(9):826-832. doi:10.1016/S0895-4356(03)00207-5.
18. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247.
19. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research 4: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(7709):1487-1490. doi:10.1136/bmj.b606.
20. Reilly BM, Evans AT. Academia and clinic. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.
21. Collins GS, Moons KGM. Comparing risk prediction models. *Br Med J*. 2012;344(may24 2):e3186-e3186. doi:10.1136/bmj.e3186.
22. Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *Bmj*. 2016;353(i3140):11. doi:10.1136/bmj.i3140.
23. Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg*. 2015;102(2):93-101. doi:10.1002/bjs.9723.
24. Krumholz HM. Why data sharing should be the expected norm. *Bmj*. 2015;350(feb05 6):h599-h599. doi:10.1136/bmj.h599.
25. Debray TPA, Riley RD, Rovers MM, Reitsma JB, Moons KGM. Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLoS Med*. 2015;12(10):1-12. doi:10.1371/journal.pmed.1001886.
26. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Assessing risk prediction models using individual participant data from multiple studies. *Am J Epidemiol*. 2014;179(5):621-632. doi:10.1093/aje/kwt298.
27. Debray TP a, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med 2013;32*. 2013;32:3158-3180. doi:10.1002/sim.5732.
28. Broeze KA, Opmeer BC, Bachmann LM, et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. *BMC Med Res Methodol*. 2009;9:22. doi:10.1186/1471-2288-9-22.
29. Ahmed I, Debray TP a, Moons KGM, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. *BMC Med Res Methodol*. 2014;14(1):3. doi:10.1186/1471-2288-14-3.
30. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines:

descriptive study. *Br Med J*. 2015;350(1088):1-9. doi:10.1136/bmj.h1088.
31. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *Bmj*. 2012;344(January):1-10. doi:10.1136/bmj.d7762.
32. Soriano JB, Lamprecht B, Ramírez AS, et al. Mortality prediction in chronic obstructive pulmonary disease comparing the GOLD 2007 and 2011 staging systems: a pooled analysis of individual patient data. *Lancet Respir Med*. 2015;3(6):443-450. doi:10.1016/S2213-2600(15)00157-5.
33. Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open*. 2012;2(6):1-10. doi:10.1136/bmjopen-2012-002152.
34. Held U, Garcia-Aymerich J, Ter Riet G, Moons KGM, Puhan MA. Missing variables in risk prediction models. *AJE (American J Epidemiol*. 2012. doi:10.1002/elan.
35. Harrell FE. *Regression Modeling Strategies*. Springer; 2015.
36. Snell KIE, Hua H, Debray TPA, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2016;69:40-50. doi:10.1016/j.jclinepi.2015.05.009.
37. Collins GS, de Groot J a, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14(1):40. doi:10.1186/1471-2288-14-40.
38. Blattenberger G, Lad F. Separating the Brier Score into Calibration and Refinement Components : A Graphical Exposition Author (s): Gail Blattenberger and Frank Lad Source : The American Statistician , Vol . 39 , No . 1 (Feb ., 1985), pp . 26-32 Published by : Taylor & Franci. *Am Stat*. 1985;39(1):26-32.
39. Murphy AH. A New Vector Partition of the Probability Score. *J Appl Metereology*. 1973;12:595-600.

CHAPTER 5: General Discussion

ACKNOWLEDGEMENTS

Acknowledgements

This dissertation was elaborated under the "Life Science" Program, in which is involved the Epidemiology, Biostatistics and Prevention Institute (EBPI) of the University of Zurich.

I would like to thank the persons who mainly supported me and my Ph.D. project.

First, I would like to thank Milo Puhan, who made me grow professionally and personally, through a privileged mentoring process. He unveiled for me the power of interdisciplinary research; he taught me how to combine technical and soft skills and gave me the possibilities to interact with some of his strict collaborators worldwide. Finally, his supervision pushed my limits towards unexpected goals, which keep surprising me and the people that I interact with.

Concerning the people that I had the honor to interact with during my Ph.D. path, I would also like to mention Leo Held for the passionate and detailed feedback that he provided me with, for instance, during the Ph.D. Committee meetings. I was also impressed by the enthusiasm for research that he conveyed to me during my first approach to EBPI, namely his course regarding statistical models for Epidemiology.

A crucial moment during my Ph.D. journey was my experience in the Netherlands with Gerben ter Riet. He gave me the possibility to obtain a flavor of the scientific approach undertaken by one of the most prestigious medical research institutes, namely the Academic Medical Center (AMC) of the University of Amsterdam. He showed me a different approach to Medicine and Research in general, providing me with a global picture that I will treasure for my whole life.

Finally, I would like to thank my Ph.D. angel, Sarah Haile, who with her expertise, passion and patience provided me with a solid and never-ending support.

Epidemiology, Biostatistics and Prevention Institute (EBPI)
University of Zurich
Switzerland

Zurich, October 8th, 2017